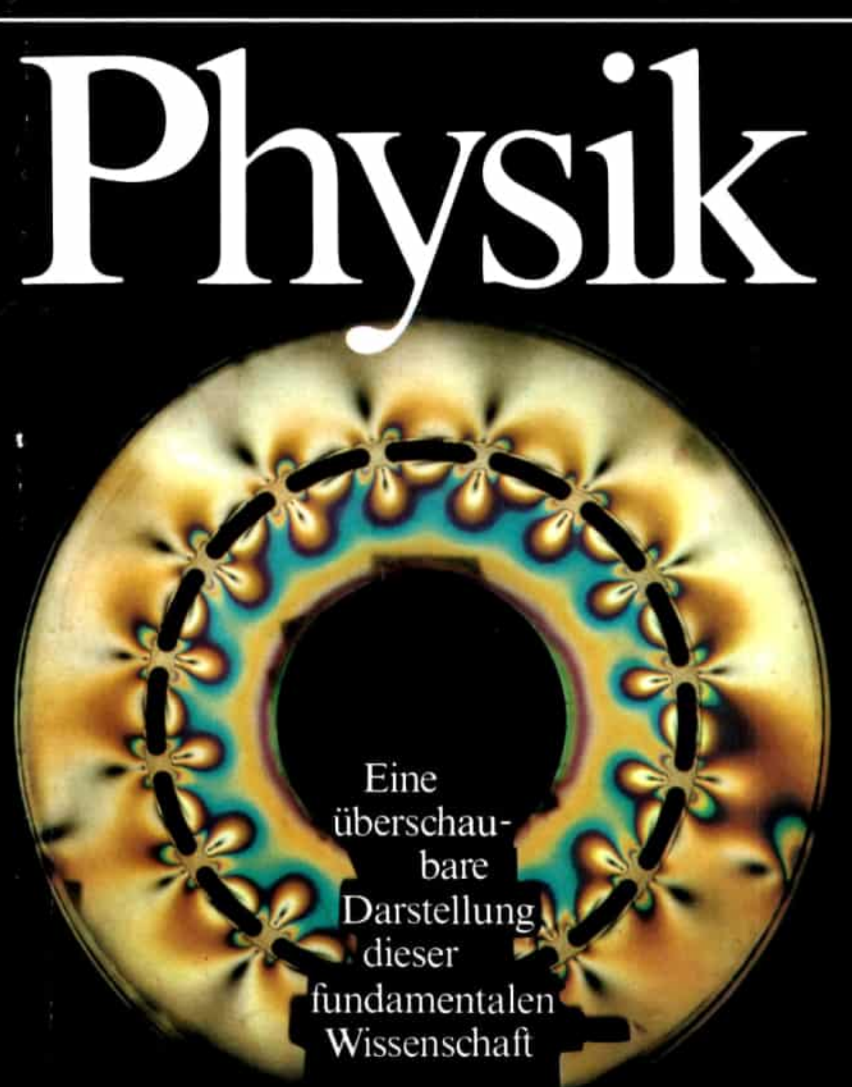
Kleine Enzyklopädie



Kleine Enzyklopädie
Physik

Kleine Enzyklopädie

Physik

Physik

Kleine Enzyklopädie Physik

Kleine Enzyklopädie

Physik

1 100 Abbildungen und

128 Tabellen sowie

32 einfarbige und

24 mehrfarbige Bildtafeln

VEB Bibliographisches Institut

Leipzig

Herausgeber
Prof. Dr. rer. nat. habil. Peter Rennert
Dr. sc. nat. Herbert Schmiedel
Prof. Dr. rer. nat. habil. Christian Weißmantel

Gutachter
Prof. Dr. Arthur Lösche

Physik / Hrsg.: Peter Rennert,
Herbert Schmiedel,
Christian Weißmantel. –
1. Aufl. – Leipzig : Bibliographisches Institut, 1986.
727, 56 S. : 1 100 Ill. (z. T. farb.)
(Kleine Enzyklopädie)
ISBN 3-323-00011-0

NE: Rennert, Peter [Hrsg.]; GT.

ISBN 3-323-00011-0

1. Auflage
© VEB Bibliographisches Institut Leipzig, 1986
Verlagslizenz-Nr. 433-130/86/86
Printed in the German Democratic Republic
Lichtsatz: INTERDRUCK, Graphischer Großbetrieb Leipzig – III/18/97
Druck und Einband: Grafischer Großbetrieb Völkerfreundschaft Dresden
Lektorat: Helmut Kahnt
Bild- und technische Redaktion: Birgit Schnabel, Monika Thiel
Herstellung: Karin Fleischer
Schutzumschlag und Einbandgestaltung: Rolf Kunze
Redaktionsschluß: Juni 1985
LSV 1107
Best.-Nr.: 577 823 8
02400

Inhaltsverzeichnis

Vorwort

Die Physik ist eine fundamentale Naturwissenschaft. Sie befaßt sich mit den elementarsten Zuständen der Materie, durch deren Wechselspiel die ungeheure Vielfalt der Natur entsteht.
Der vorliegende Band „Kleine Enzyklopädie Physik" soll dem interessierten Leser einen Überblick über die wichtigsten physikalischen Gesetze und deren Anwendung geben. Physikalische Prinzipe spielen in allen Naturwissenschaften eine wichtige Rolle und sind Grundlage der Technik und Technologie.
Die „Kleine Enzyklopädie Physik" ist ein populärwissenschaftliches Buch, das sich an einen breiten Leserkreis wendet. Sowohl der wissensdurstige Laie, der Lernende, als auch der Lehrende, der Fachmann, sollen angesprochen werden. So haben sich die Autoren dieser Darstellung des Gesamtgebietes der Physik bemüht, Leichtverständlichkeit mit Exaktheit zu verknüpfen.
Die Physik ist eine im wesentlichen experimentelle Wissenschaft, sie benutzt jedoch die Mathematik als ein äußerst wichtiges Hilfsmittel. Im vorliegenden Buch wurden deshalb je ein gesonderter Abschnitt den in der Physik häufig verwendeten mathematischen Methoden als auch den modernen Meßmethoden der Physik gewidmet.

Der begrenzte Umfang des Buches zwang die Herausgeber zu einer sorgfältigen Auswahl des Stoffes und der Beschränkung auf die wichtigsten Sachverhalte der betreffenden Teilgebiete der Physik. Als Autoren der einzelnen Abschnitte des vorliegenden Bandes der populären Reihe „Kleine Enzyklopädie" des VEB Bibliographischen Institutes Leipzig konnten erfahrene Wissenschaftler gewonnen werden, die durch gewissenhafte Aufbereitung des Stoffes sowohl die klassischen als auch die modernen Erkenntnisse der Physik übersichtlich und knapp erläutert haben. So findet man in der „Kleinen Enzyklopädie Physik" nicht nur die wichtigsten Informationen zum Inhalt solcher Standardgebiete der Physik, wie Mechanik, Elektrodynamik, Quantenmechanik, Relativitätstheorie, Hochenergiephysik, Thermodynamik und Statistische Physik, sondern auch Einführungen in die Biophysik und die Physikalische Chemie, die Grundlagen der Astro- und Weltraumphysik und der Geophysik einschließlich der Energie- und Umweltproblematik. Der Abschnitt „Angewandte Elektrizitätslehre" enthält die physikalischen Grundlagen der Elektronik, dem Herzstück der wissenschaftlich-technischen Revolution.

Leipzig, Sommer 1985 — Herausgeber und Verlag

Autoren

Ammer, Christian, Dipl.-Phys. 6.10.
Berg, Gunnar, Dr.-Ing. Dr. rer. nat. 2.2., 6.5., 6.6., 6.7.
Brauer, Karl-Heinz, Dr. rer. nat. habil. 2.4., 2.5.
Breuer, Klaus, Prof. Dr. sc. nat. 5.2.–5.11.
Brühl, Hans-Gert, Dr. rer. nat. 7.8.
Conrad, Walter, 1.
Frieser, Alfred, Dr. rer. nat. 7.3.–7.3.9., 7.4.
Hänsel, Christian, Prof. Dr. sc. nat. 10.
Hart, Hans, Prof. Dr. rer. nat. habil. 7.2.
Herlt, Eduard, Dr. sc. nat. 3.1.
Klaua, Manfred, Dr. 6.10.
Kleinschmidt, Jürgen, Dr. rer. nat. 4.11.
Kleint, Christian, Dr. sc. 7.1.
Knoll, Helmut, Dr. 9.
Kramer, Dietrich, Dr. rer. nat. habil. 3.3.
Kreher, Konrad, Dr. sc. nat. 6.3.8.
Krüger, Horst, Dr. rer. nat. 6.9.
Kühn, Günther, Dr. sc. nat. 6.2.
Limmer, Stefan, Dr. rer. nat. 11.
Meier, Wolfgang, Dr. rer. nat. 3.5.
Neumann, Hans, Dr. sc. nat. 7.3.9.–7.3.12.
Paufler, Peter, Prof. Dr. sc. nat. 6.1.
Przyborowski, Frank, Dr. rer nat. 7.7.
Rennert, Peter, Prof. Dr. rer. nat. habil. 2.1., 2.2., 2.3., 2.6., 2.7., 4.1.–4.7., 6.3., 6.4.
Salié, Nikolaus, Dr. sc. nat. 3.1., 3.4.
Schmidt, Werner, Prof. Dr. sc. 8.
Schmiedel, Herbert, Doz. Dr. sc. nat. 7.4.
Schmutzer, Ernst, Prof. Dr. sc. nat. 3.2., 3.6.
Wagner, Norbert, Dr. rer. nat. 6.8.
Wenzel, Horst, Prof. Dr. rer. nat. habil. 13.
Wieczorek, Eberhard, Dr. sc. nat. 5.1.
Wiederhold, Gerhard, Dr. sc. nat. 7.5., 7.6.
Wilhelmi, Bernd, Prof. Dr. sc. nat. 4.11.
Zimmermann, Helmut, Prof. Dr. rer. nat. habil. 12.

Quellennachweis für Abbildungen

Wissenschaftliche Aufnahmen und Reproduktionen stellten zur Verfügung:

ADN-Zentralbild, Berlin · Akademie der Wissenschaften der DDR, Redaktion „Spektrum" · Akademie Verlag, Berlin, Redaktion „Wissenschaft und Fortschritt" · BI-Archiv, Leipzig · Cristaltec, Forschungszentrum des CNRS, Grenoble · Deutsche Fotothek, Dresden · VEB Deutscher Verlag für Grundstoffindustrie, Leipzig · Deutsches Museum, München · DEWAG Berlin · VEB Fachbuchverlag, Leipzig · Dr. rer. nat. Helmut Giegengack, Karl-Marx-Stadt · Dr. Haberland, Leipzig · VEB Hochvakuum Dresden · Karl-Schwarzschild-Observatorium, Tautenburg · Karl-Sudhoff-Institut, Leipzig · Dr. K. W. Keller, Halle · Dr. Manfred Klaua, Halle · Konrad Kopsch, Niemegk · Kwasan Observatory, Kyoto · Prof. Dr. E. W. Müller (verst.) · NASA, California Institute of Technology, Pasadena, California · Dipl.-Phys. Dietrich Pätzold, Berlin · Dr. rer. nat. Przyborowski, Neukirchen · Lothar Richter, Jena · Fred Schindler, Leipzig · Siemens-Presseinformation, München · Urania-Verlag, Leipzig, Jena, Berlin · VEB Verlag Volk und Wissen, Berlin · Dr. sc. nat. Gerhard Wiederhold, Jena · VEB Carl Zeiss Jena, Redaktion Jenaer Rundschau

Die Textabbildungen wurden nach Angaben der Autoren angefertigt von:

Gerhard Pippig, Großdeuben.

Die Farbtafeln 34, 50, 51 zeichnete:

Joachim Zindler, Leipzig

1. Geschichtlicher Überblick

1.1. Anfänge physikalischer Beobachtungen und Vorstellungen

Der Weg der wissenschaftlich-technischen Entwicklung begann mit Beobachtungen und Erfahrungen. Sie sammelten sich während der Jahrtausende und wurden von Generation zu Generation überliefert. Stimuliert wurde die Entwicklung immer wieder durch Bedürfnisse der Gesellschaft – sei es, daß es Bauten zu errichten, Land zu vermessen und zu bewässern galt, sei es, daß man nach Möglichkeiten suchte, sich auf Handelsreisen zu orientieren.

Allmählich schälten sich durch die vielfältige Praxis verallgemeinernde, zunächst vorwiegend mechanische Begriffe, wie Masse, Kraft und Bewegung, heraus. Gemeinsamkeiten unterschiedlicher Vorgänge und Erscheinungen wurden beobachtet und, sofern sie sich als nützlich erwiesen, häufig auch wieder verwertet. So fand man etwa, daß Reibung stets Wärme entstehen läßt und zusätzlichen Kraftaufwand erforderlich macht. Das führte einerseits zu der wichtigen Erfindung der ersten „Feuerzeuge", andererseits lernte man, Reibung und Kraftaufwand durch geeignete Hilfsmittel zu vermindern. Man bediente sich mit Geschick des Hebels, des Keils und der Rolle, ohne zunächst zu untersuchen, worauf ihre Wirkung eigentlich beruhte; man wendete statische, dynamische, hydraulische Gesetze an, die erst viel später formuliert wurden usw. Entdeckungen, mit denen sich in der Praxis nichts anfangen ließ, z.B. die elektrostatische und die magnetische Anziehung, blieben vorerst unbeachtet.

Die Frage nach dem „Warum?" der Erscheinungen, nach Aufbau, Entstehung und Zusammenhang der Welt wurde vereinzelt schon in den orientalischen Sklavenhalterstaaten, vor allem aber von den griechischen Naturphilosophen gestellt, wie überhaupt die Anfänge der gesamten Naturwissenschaft in die Philosophie eingebettet sind. Zielgerichtete Experimente, „Fragen an die Natur", gab es noch nicht. Jedoch schlugen sich die Resultate vereinzelter Untersuchungen in den Ideen und Werken der Naturphilosophen nieder. Dabei kam es u. a. zu so bemerkenswerten Leistungen wie der materialistischen Atomistik LEUKIPPS (um 450 v. u. Z.) und DEMOKRITS (um 460 bis um 370 v. u. Z.), die sich die Welt aus unerschaffbaren, unzerstörbaren, in ständiger Bewegung befindlichen und miteinander in Wechselwirkung stehenden Elementarteilchen zusammengesetzt dachten.

Unter den Philosophen jener frühen Epoche nimmt ARISTOTELES (384–322 v. u. Z.) eine herausragende Stellung ein, zum einen wegen seiner geradezu enzyklopädischen Leistung, zum anderen wegen des nachhaltigen Wirkens seiner Lehren und Ideen. Er systematisierte das Wissen seiner Zeit, gliederte die Wissenschaft nach Disziplinen und prägte u. a. die Bezeichnung Physik, die er allerdings noch in einem allgemein naturphilosophischen Sinne verstand. Seine physikalischen Vorstellungen im engeren Sinne waren spekulativ und führten zu Irrtümern und Trugschlüssen. So ergab sich aus seinen Gedankengängen über „natürliche" und durch „Beweger" erzwungene Bewegungen nicht nur die Schlußfolgerung einer gewichtsabhängigen Fallgeschwindigkeit (eine Behauptung, die sich durch einfachste Experimente hätte widerlegen lassen), sondern auch eine scheinbare Begründung für die Unmöglichkeit eines „leeren" Raumes.

ARCHIMEDES (287–212 v. u. Z.) gilt als einer der ersten großen „Mechaniker" oder „Ingenieure", die Naturgesetze und Berechnungen bewußt auf die Lösung praktisch-technischer Aufgaben anwendeten. Seine Leistungen auf den Gebieten der Mathematik, Mechanik, Hydrostatik und Technik nötigen uns heute noch Bewunderung ab. Unter seinen Hauptwerken – nicht alle blieben erhalten – waren Abhandlungen „Über schwimmende Körper", „Über das Gleichgewicht ebener Flächen und über den Schwerpunkt ebener Flächen", „Über den Brennspiegel", „Über die Anfertigung von Sphären". ARCHIMEDES berechnete Inhalt und Oberfläche zahlreicher geometrischer Körper, wobei er teilweise Verfahren vorwegnahm, die viele Jahrhunderte später zur Infinitesimalrechnung ausgebaut wurden. Er leitete das Hebelgesetz ab und fand das nach ihm benannte Prinzip für den Gewichtsverlust von Körpern in Flüssigkeiten. Er setzte, z. B. im Schiffsbau, die schiefe Ebene, Winden, Rollen und Flaschenzüge ein, erfand die „Archimedische Schraube" zum Wasserfördern, konstruierte ein einfaches Planetarium und ersann Schleudern u. a. Kriegsgerät, die bei der Verteidigung seiner Heimatstadt Syrakus gegen die Römer eingesetzt wurden. Obwohl als sicher gilt, daß ARCHIMEDES bisweilen für seine Forschungen Modelle anfertigen ließ und benutzte, war er doch kein Experimentalphysiker, der genau umrissenen Fragestellungen durch eine Reihe von Experimenten nachging. Das gilt auch für seine Nachfolger.

Der im 2. Jh. v. u. Z. lebende KTESIBIOS arbeitete vor allem auf pneumatischem und hydraulischem Gebiet. Man schreibt ihm die Erfindung einer Wasseruhr mit Zeigerwerk zu, ferner die Konstruktion einer „Wasserorgel", eines Druckluftgeschützes sowie die Erfindung von doppeltwirkenden Druckpumpen und Feuerspritzen.

Sein Schüler HERON von Alexandria (um 100 v. u. Z.) wurde als Mathematiker, Geodät und Techniker bekannt. Von ihm stammt die Formel für die Berechnung der Dreiecksfläche aus den drei Seiten. Er erfand ein Visierinstrument für Messungen im Gelände, einen Wegmesser für Wagen (er enthielt das erste bekanntgewordene Untersetzungsgetriebe), ein Thermoskop, den Heronsball und mit der Äolipile einen auf Rückstoßwirkung beruhenden Vorläufer der Dampfturbine. Weder diese noch verschiedene Automaten u. a. Erfindungen HERONS erlangten in der kaum auf Entwicklung der Produktivkräfte bedachten Sklavenhaltergesellschaft praktische Bedeutung. Sie blieben technische Spielereien, die allenfalls zu Kultzwecken (z. B. eine Vorrichtung zum selbsttätigen Öffnen von Tempeltüren und ein Weihwasserautomat) angewandt wurden.

1.2. Zwischen Antike und Renaissance

Im römischen Imperium begnügte man sich im wesentlichen mit der ingenieurtechnischen Nutzung übernomme-

ner Kenntnisse. Immerhin wurden von römischen Autoren mehrere umfangreiche Sammelwerke über wissenschaftliche und technische Gegenstände verfaßt. Zu ihnen zählt z.B. die 37 Bücher umfassende „Naturalis historia“ des Gajus Plinius SECUNDUS (des Älteren, 23–79 u.Z.) ebenso wie das Werk „De architectura“ von Vitruvius POLLIO (geb. um 70 v. u. Z.), das nicht nur die Baukunst und Bautechnik, sondern auch Uhren u. a. technische Geräte und Einrichtungen behandelt.
Der in Alexandria wirkende Claudius PTOLEMÄUS (etwa 90–160) faßte in seinem „Almagest“ das astronomische Wissen zusammen und führte das geozentrische Weltbild weiter. Er untersuchte die **Lichtbrechung** und die atmosphärische Strahlenbrechung und legte die Resultate in Tabellen nieder, ohne allerdings das Brechungsgesetz selbst zu erkennen.
Um die Bewahrung der antiken Wissenschaft erwarben sich besonders die Araber Verdienste, deren Einflußbereich sich in der zweiten Hälfte des ersten Jahrtausends zeitweise von Spanien über Mittelasien bis an die Grenzen Chinas und Indiens erstreckte. In zahlreichen Städten, wie Bagdad, Buchara, Samarkand, Cordoba, entstanden Hochschulen, Bibliotheken und wissenschaftliche Zentren. Dort wurden die Schriften der Antike ins Arabische übersetzt, kommentiert und von Gelehrten auch zur Grundlage weiterführender Forschungen gemacht. So untersuchte ALHAZEN (Abu Ali al-Hasan ibn al-HASAN ibn-HAITAM, um 965–um1038) die Lichtbrechung. Er versuchte, aus der Dauer der Dämmerung die Höhe der Atmosphäre zu errechnen und konzipierte eine Theorie des Sehens. Seine Vorstellungen und Resultate sind in dem Werk „Optik“ niedergelegt.
Der usbekische Gelehrte Abur-Reihan Muhamed ibn-Achmed al BIRUNI (973–1048) hinterließ ein umfangreiches wissenschaftliches Sammelwerk. Er forderte die Trennung von Religion und Wissenschaft und betonte die überragende Bedeutung von Beobachtungen, Fakten und auch Experimenten für die Forschung. Weit über die Grenzen der arabischen Welt wurden AVICENNA (Abu Ali Husain ibn SINA, 980–1037) bekannt. Mehr als 200 Schriften von seiner Hand wurden bisher aufgefunden. Sie befassen sich, ebenso wie eine 18bändige Enzyklopädie, mit Philosophie, Mathematik, Geologie, Geografie, Physik, Mineralogie. AVICENNAS „Kanon der Medizin“ galt für Jahrhunderte auch in Europa als maßgebliches medizinisches Handbuch.
Mit dem Aufkommen des Feudalismus und der Ausbreitung der christlichen Religion in Mittel- und Westeuropa wurde die Scholastik zur beherrschenden Philosophie. Das geistige Leben blieb im wesentlichen auf die Klöster beschränkt, Aufgabe der Wissenschaft war es in erster Linie, der Theologie zu dienen, ihre Dogmen zu „belegen“, wobei man „geeignete“ Teile der alten Philosophie, z. B. Gedankengänge des ARISTOTELES oder des PTOLEMÄUS, geschickt mit den christlichen Lehren zu verbinden suchte. Fortschrittliche Gelehrte und ihre Ideen wurden totgeschwiegen oder zum Schweigen gebracht. So durfte Roger BACON (um 1214–1294) wegen seines Auftretens gegen scholastischen Dogmatismus nicht mehr lehren und wurde mehrfach, schließlich bis zu seinem Tode, eingekerkert. Er hatte u. a. über Optik geschrieben (Regenbogen, Luftspiegelungen, optische Lesehilfen) und hatte Quellenstudium, Erfahrung, Beobachtungen und Experimente als Grundlage naturwissenschaftlicher Forschung gefordert.

1.3. Die Physik auf dem Wege zur selbständigen Wissenschaft

Trotz solcher wissenschaftsfeindlicher Haltung kam im Mittelalter der physikalisch-technische Fortschritt nicht völlig zum Stillstand. Der Kompaß, unentbehrliche Navigationshilfe und Voraussetzung für Schiffahrt auf hoher See, wurde – aus China übernommen – in Europa bekannt. Mechanische Turmuhren wurden konstruiert.
Im 11. und vor allem im 12. Jh. wurden zahlreiche klassische Arbeiten aus dem Arabischen oder Griechischen in die lateinische Sprache übersetzt und, immer wieder abgeschrieben, verbreitet.
Auch europäische Gelehrte meldeten sich zu Wort: 1235 wurde erstmals das Perpetuum mobile in einer Abhandlung beschrieben. 1269 erschien der „Brief über den Magneten“ von Pierre de MARICOURT. 1283 tauchten in Italien Berichte über (wahrscheinlich schon vorher erfundene) Brillen mit geschliffenen Gläsern auf. BACON vermutete eine zum Erdmittelpunkt gerichtete Anziehungskraft als Ursache des Gewichtes. Um 1340 legte Konrad von MEGENBERG (um 1309–1374) das „Buch der Natur“ vor, die erste Naturgeschichte in deutscher Sprache. In mehreren europäischen Städten wurden Universitäten gegründet, z. B. Bologna 1119, Prag 1348, Heidelberg 1386, Leipzig 1409. Von Italien ausgehend entstanden in West- und Mitteleuropa Vorbedingungen für frühkapitalistische Produktionsverhältnisse. Die Renaissance brachte einen bis dahin einmaligen Aufschwung von Kultur, Wissenschaft und Technik.
Um 1435 begann Johann GUTENBERG (1399–1468) mit Arbeiten an wiederverwendbaren Metallettern. 1492 erreichte Christoph KOLUMBUS (1451–1506) Guanahani, Kuba und Haïti.
1543 erschien die revolutionär wirkende Arbeit des polnischen Domherren und Astronomen Nikolaus KOPERNIKUS (1473–1543) „Über die Umläufe der Himmelskörper“. Sie widerlegte endgültig das von der Scholastik allgemein anerkannte Weltbild einer im Zentrum des Kosmos ruhenden Erde.
Johannes KEPLER (1571–1630) vollendete das Werk des KOPERNIKUS, indem er Ellipsen an die Stelle kreisförmiger Planetenbahnen setzte und die nach ihm benannten Gesetze formulierte. 1611 beschrieb er das astronomische Fernrohr. Auch mit der Konstruktion von Rechenmaschinen befaßte er sich.
Einer der größten Künstler, Techniker und Wissenschaftler der Renaissance war Leonardo da VINCI (1452–1519). Allein seine Ideen und Vorschläge auf technischem Gebiet wirkten für Jahrhunderte fruchtbringend. Er beschäftigte sich auch mit Stoß- und Bewegungsvorgängen, untersuchte die Gleit- und die Rollreibung, entwickelte Vorstellungen über das Zustandekommen des Himmelsblaus, über den Vogelflug und viele andere Gegenstände.
Vorschläge für das Perpetuum mobile (und Hinweise auf seine Unausführbarkeit) finden sich gleichfalls unter seinen Notizen.
Der niederländische Baumeister und Ingenieur Simon STEVIN (1548–1620) untersuchte statische Probleme, die Zusammensetzung und Zerlegung von Kräften (Kräfteparallelogramm) und die schiefe Ebene.
William GILBERT (1544–1603) führte Untersuchungen auf dem Gebiet der Elektrostatik, vor allem aber auf dem des Magnetismus durch, betrachtete die Erde als großen Magneten und benutzte für Experimente und Messungen ein „Erdmodell“. Er unterschied zwischen durch Reiben auf-

ladbaren und nicht aufladbaren Körpern und prägte den Begriff „elektrisch".
Mikroskop und Fernrohr wurden um 1600 von niederländischen Brillenmachern und Linsenschleifern erfunden und später u. a. von Robert HOOKE (1635–1703) wesentlich verbessert. Galileo GALILEI (1564–1642) war der erste überragende Physiker, der Experiment und Berechnung konsequent zur Gewinnung physikalischer Erkenntnisse nutzte, von präzisen Fragestellungen ausging und sich auf deren Kern konzentrierte, während er Nebeneffekte (z. B. den Luftwiderstand beim Wurf) zunächst unberücksichtigt ließ. Er klärte die Begriffe Geschwindigkeit und Beschleunigung, leitete die Fall- und Wurfgesetze ab, untersuchte Pendelschwingungen, arbeitete über Festigkeitslehre und fand Ansätze des Trägheitsgesetzes. Anknüpfend an Berichte aus den Niederlanden konstruierte GALILEI Fernrohre, mit denen er Jupitermonde und die Phasen der Venus entdeckte, Mondkrater und Sonnenflecken beobachtete. Durch seine Entdeckungen und Beobachtungen wurde er zum Verfechter des kopernikanischen Weltsystems. In seinem Werk „Dialog über die zwei hauptsächlichen Weltsysteme ..." (1632) stellte er es überzeugend dar. Wegen seiner von der offiziellen kirchlichen Lehrmeinung abweichenden Darlegungen wurde GALILEI vor das Inquisitionsgericht geladen und zum Abschwören seiner Lehren gezwungen. Bis an sein Lebensende nahezu unter ständigem Hausarrest stehend und fast erblindet schrieb er die ebenfalls in Gesprächsform gehaltenen „Unterredungen und mathematischen Demonstrationen über zwei neue Wissenszweige, die Mechanik und die Fallgesetze betreffend". Sie erschienen 1638 in Holland.
GALILEIS Schüler Evangelista TORRICELLI (1608–1647) führte, ausgehend von der begrenzten Höhe einer durch Kolbenpumpen förderbaren Wassersäule, Untersuchungen durch, in deren Verlauf er den Luftdruck und Luftdruckschwankungen nachwies, deutete und 1643 das Quecksilberbarometer erfand. Die alte Vorstellung von der Unmöglichkeit eines leeren Raumes („horror vacui") wurde durch seine Arbeiten widerlegt, zumal 1648 Blaise PASCAL (1623–1662) einen weiteren Beweis beisteuerte, indem er die Abnahme des zu gleichen Zeiten festgestellten Luftdruckes mit der Höhe zeigte. Damit war zugleich die Voraussetzung für barometrische Höhenmessungen geschaffen. Nahezu gleichzeitig mit TORRICELLI und PASCAL wandte sich der Magdeburger Ratsherr, Bürgermeister und Physiker Otto von GUERICKE (1602–1686) der „Leere" zu. Nachdem er anfänglich Gefäße mit Hilfe von Wasserpumpen und Handspritzen zu evakuieren versucht hatte, konstruierte er später wirkungsvolle Luftpumpen. Mit ihrer Hilfe führte er die eindrucksvollen Demonstrationen durch, die seinen Namen in der wissenschaftlichen Welt bekannt werden ließen. Er beobachtete die Wägbarkeit und Elastizität der Luft, ihr wärmeabhängiges Ausdehnungsvermögen, ihre Rolle für die Schallausbreitung und bei Verbrennungsvorgängen. An seinem Hause brachte er ein Wasserbarometer an, mit dem er meteorologische Beobachtungen anstellte und sogar Voraussagen wagte. Eine einfache Reibungselektrisiermaschine, bestehend aus einer drehbaren und mit der Hand geriebenen Schwefelkugel, geht gleichfalls auf GUERICKE zurück.
Die Lichtbrechung hatte neben anderen bereits Claudius PTOLEMÄUS untersucht; jedoch wurde das Brechungsgesetz erst um 1620 von Willebrord SNELL (Snellius van ROJEN, 1591–1626) erkannt.
Lichtbeugung und Interferenz untersuchte Francesco Maria GRIMALDI (1618–1663). Seine Erklärungsversuche kamen der Vorstellung sich ausbreitender Lichtwellen nahe.
Die umstrittene Frage, ob die Lichtgeschwindigkeit unendlich oder endlich groß sei, entschied der dänische Astronom Ole Christensen RÖMER (1644–1710), indem er die Lichtgeschwindigkeit aus scheinbaren Schwankungen der Umlaufzeit des Jupitermondes „Jo" bestimmte.
In seiner „Abhandlung über das Licht" gab Christian HUYGENS (1629–1695) die erste geschlossene Darstellung der Lichtwellenhypothese. Er hielt Licht für Längsschwingungen, die sich schallähnlich, nach Gesetzen der Mechanik, im „Äther" fortpflanzen. Ihn dachte sich HUYGENS aus unwägbaren, kleinsten, nahezu ideal elastischen und zugleich sehr harten Teilchen bestehend. Erscheinungen wie geradlinige Ausbreitung, Reflexion, Beugung ließen sich so leicht erklären. Bekannte Gelehrte – unter ihnen Leonhard EULER (1707–1783) und Michail Wassiljewitsch LOMONOSSOW (1711–1765) – stimmten HUYGENS zu. HUYGENS untersuchte die kurz zuvor entdeckte Doppelbrechung, arbeitete über Stoßgesetze, die Zentrifugalkraft und Pendelschwingungen. Er verbesserte astronomische Fernrohre, erfand die Federuhr mit Unruh und die Pendeluhr.

1.4. Physik im Zeichen der Mechanik

Im Laufe des 17. Jh. war das Experiment zur wichtigsten Arbeitsmethode des Physikers geworden. Unterstützt wurde sie durch Hilfsmittel wie Mikroskop, genaue Uhren und empfindliche Waagen.
Algebra und analytische Geometrie erleichterten die Auswertung und Interpretation von Beobachtungen und Meßwerten, Logarithmen verkürzten langwierige Rechenarbeit; später wurden Differential- und Integralrechnung – besonders von Gottfried Wilhelm LEIBNIZ (1646–1716) und Isaac NEWTON (1643–1727) entwickelt – zu einem vielseitig anwendbaren Werkzeug. Eine Fülle von Beobachtungen, Fakten und Erkenntnissen lag vor, an die Naturwissenschaftler anknüpfen konnten.
NEWTON nutzte dieses Material. Besonders durch seine Leistungen auf dem Gebiet der Mechanik begründete er auf der Basis eines absoluten Raumes und einer absoluten Zeit die klassische Mechanik. Er entdeckte die Gravitation, formulierte das Gravitationsgesetz und leitete daraus die Keplerschen Gesetze der Planetenbewegung mathematisch her. In seinem Hauptwerk „Mathematische Prinzipien der Naturlehre" (1687) werden wichtige Grundbegriffe der Mechanik, wie Masse, Gewicht, Kraft, geklärt. Die Dynamik wird mit Hilfe der Newtonschen Axiome abgehandelt. Auch auf Schwingungen und das Verhalten bewegter Flüssigkeiten geht NEWTON ein.
Die Newtonsche Mechanik wurde von Zeitgenossen und Nachfolgern ausgebaut; LEIBNIZ präzisierte den Begriff der „lebendigen Kraft" (kinetische Energie); Daniel BERNOULLI (1700–1782) leistete wichtige Beiträge zur Hydrodynamik; Joseph-Louis LAGRANGE (1736–1813) arbeitete auf dem Gebiet der analytischen Mechanik, während Pierre-Simon LAPLACE (1749–1827) in seiner fünfbändigen „Himmelsmechanik" die Newtonsche Mechanik auf Vorgänge im gesamten Weltall ausdehnte. Seit 1666 befaßte sich NEWTON mit optischen Arbeiten. Er untersuchte die Zerlegung weißen Lichtes durch Prismen und seine Wiedervereinigung, Brechungs- und Beugungserscheinungen und Körperfarben. 1668 konstruierte er das Spiegelteleskop. NEWTONS Werk „Optik" erschien 1704 nach vorausgegangenen Einzelveröf-

fentlichungen. Im Gegensatz zu HUYGENS ging NEWTON von einer Emissionshypothese aus, die sich Licht als einen Strom schnellfliegender Teilchen vorstellte. Zahlreiche optische Erscheinungen ließen sich damit erklären, andere (z. B. Beugung und Interferenz) nur durch Zusatzannahmen, mit denen sich NEWTON in gewisser Weise der Wellenhypothese näherte. Für viele Jahrzehnte wurde NEWTONS Emissionstheorie von der großen Mehrzahl der Wissenschaftler anerkannt. Die Entscheidung für die Wellentheorie des Lichtes im Rahmen der klassischen Physik fiel in den ersten Jahrzehnten des 19. Jh. Zu ihr trugen bei: Arbeiten von Thomas YOUNG (1773–1829), der sich besonders den Interferenzerscheinungen zuwandte und aus ihnen die Lichtwellenlängen errechnete; Untersuchungen von Etienne-Louis MALUS (1755–1812), der die Polarisation erforschte und durch die Annahme transversaler Schwingungen zu erklären suchte; vor allem aber die Arbeiten von Augustin-Jean FRESNEL (1788–1827), der zeigte, daß Interferenz bei polarisiertem Licht nur auftreten kann, wenn die Schwingungsebenen der interferierenden Strahlen einander parallele Komponenten haben. Zusätzliche Beweise dafür, daß es sich bei Licht um „Ätherschwingungen" handelt, wurden um die Mitte des 19. Jh. durch die Messung der Lichtgeschwindigkeit in verschiedenen Medien erbracht.
Arbeiten auf dem Gebiet der Elektrizitätslehre waren im 17. und 18. Jh. durch die Elektrostatik gekennzeichnet. Der englische Physiker Stephan CRAY (1670–1736) entdeckte um 1727, daß Elektrizität sich durch manche Stoffe fortleiten ließ und durch andere nicht. 1746 wurde Elektrizität durch Eisendrähte auf etwa 4 km weitergeleitet, wobei die damals unmeßbar hohe Übertragungsgeschwindigkeit überraschte. Um 1730 unterschied Charles Francois de Cisternay DUFAY (1698–1738) zwischen Glas- und Harzelektrizität (Georg Christoph LICHTENBERG, 1742–1799, führte dafür die Bezeichnungen „+" und „−" ein).
Elektrisiermaschinen, deren Schwefelkugel bald nach GUERICKE durch eine Glaskugel oder einen Glaszylinder ersetzt worden war, erhielten um 1745 ein Reibzeug zur Elektrizitätsabnahme (Johann Heinrich WINKLER, 1703–1770); der Schweizer Martin von PLANTA (1727–1772) gab der Reibungselektrisiermaschine die verbreitete Form mit drehbarer Glasplatte.
Um 1745 wurde die Leidener Flasche von Georg Ewald von KLEIST (1700–1748) und Pieter van MUSSCHENBROEK (1692–1761) erfunden.
Der Luftelektrizität wendete man sich verstärkt um 1750 zu, Benjamin FRANKLIN (1706–1790) u.a. wiesen den Blitz als elektrische Erscheinung nach; 1752 wurde der Blitzableiter erfunden. 1753 legte LOMONOSSOW eine Arbeit über Luftelektrizität vor, in der er die Elektrizität auf eine Bewegung des Äthers zurückführte.
Mit der Drehwaage (1785) ersann Charles Augustin de COULOMB (1736–1806) ein empfindliches Meßinstrument zur Bestimmung der elektrischen (und magnetischen) Anziehung (Coulombsches Gesetz) und auch zur Messung der Gravitationskonstante. Wirkungen „tierischer" Elektrizität waren seit dem Altertum bekannt. Der Arzt und Physiker Luigi GALVANI (1737–1798) sah im Zucken von Froschschenkeln, die über Stücke unterschiedlicher Leiter miteinander in Verbindung standen oder sich nahe einer in Tätigkeit befindlichen Elektrisiermaschine befanden (1780), gleichfalls eine Äußerung tierischer Elektrizität und vertrat diese (irrige) Ansicht hartnäckig. Alessandro VOLTA (1745–1827), seit 1774 mit Forschungen auf dem Gebiet der Elektrizität befaßt, erdachte und konstruierte (unabhängig von anderen) den Elektrophor, verbesserte das Elektroskop zur Anzeige und Abschätzung elektrischer Ladungen und wandte sich den Arbeiten von GALVANI zu, dessen Schlußfolgerungen er als falsch erkannte (um 1794). Er erklärte sie als Berührungselektrizität und stellte, wie zahlreiche Zeitgenossen und Nachfolger, „galvanische" Elemente und Batterien (Voltasche Säule, 1800) zusammen. Sie machten über längere Zeit gleichmäßig fließende elektrische Ströme verfügbar.
Damit begann eine neue und für die Folgezeit wichtigste Etappe der Elektrizitätslehre, Voraussetzung der Elektrotechnik. Sehr bald fand man die Wärmewirkungen elektrischer Ströme, 1822 die Thermoelektrizität (Thomas Johann SEEBECK, 1770–1831). Anthony CARLISLE (1768–1804) und William NICHOLSONS (1745–1815) beobachteten 1800 die „Zersetzung" des Wassers durch elektrischen Strom und begründeten damit die Elektrochemie. Ihre Arbeiten wurden von Johann Wilhelm RITTER (1776–1810) fortgeführt, der auch einen Vorläufer des Akkumulators angab.
Wassili Wladimirowitsch PETROW (1761–1834) entdeckte 1802 den elektrischen Lichtbogen, den wenig später auch Humphry DAVY (1778–1829) fand und nutzte. Mit Hilfe der Schmelzflußelektrolyse isolierte er u. a. die Elemente Magnesium, Kalium und Natrium.
Hans Christian OERSTED (1777–1851) hatte bereits mit RITTER nach Zusammenhängen zwischen Elektrizität und Magnetismus gesucht. 1820 entdeckte er das Magnetfeld des elektrischen Stromes (s. u.)
Georg Simon OHM (1787–1854) untersuchte quantitative Beziehungen zwischen Grundgrößen des elektrischen Stromes und veröffentlichte 1827 in seiner Arbeit „Die galvanische Kette, mathematisch behandelt" das nach ihm benannte Gesetz.
Zu Beginn des 18. Jh. traten für Messungen geeignete Thermometer an die Stelle der zwar seit langem bekannten, jedoch nur für Vergleiche oder Schätzungen brauchbaren Thermoskope. Gabriel Daniel FAHRENHEIT (1686–1736) schuf Alkohol-, um 1714 Quecksilberthermometer. Die von ihm vorgeschlagene Temperaturskale ist in einigen Ländern noch heute in Gebrauch. Auch unternahm er Siedepunktsuntersuchungen, vor allem über Zusammenhänge zwischen Siedepunkt und Luftdruck. Der schwedische Physiker Anders CELSIUS (1701–1744) führte 1742 die heute übliche Temperaturskale ein. Um die Jahrhundertmitte wurden Begriffe wie spezifische Wärme(kapazität) und latente Wärme geprägt (Joseph BLACK, 1728–1799, Henry CAVENDISH, 1731–1810).
Wesentliche Impulse zur Beschäftigung mit der Wärme entsprangen der am Beginn der industriellen Revolution geschaffenen Dampfmaschine (Verbreitung etwa seit 1775). Zahlreiche Wissenschaftler, Techniker, Unternehmer interessierten sich nunmehr für die „aus Wärme erzeugbare Kraft". Die 1824 veröffentlichte Arbeit von Nicolas Léonard Sadi CARNOT (1796–1832), „Überlegungen zur bewegenden Kraft des Feuers ...", ist noch heute bedeutsam für die Konstruktion von Wärmekraftmaschinen und die Möglichkeiten und Grenzen ihrer Nutzung. Über das Wesen der Wärme existierten gleichfalls zwei einander widersprechende Ansichten. Die einen sahen in der Wärme eine Bewegungsart. Zu ihnen zählten HUYGENS, EULER, LOMONOSSOW, Antoine-Laurent LAVOISIER (1743–1794). Die Mehrzahl der Wissenschaftler jedoch trat für einen unsichtbaren, unwägbaren Wärmestoff ein, der sich in Körpern anhäufen oder zwischen Körpern übergehen könne. Mit dieser Annahme ließen sich viele Grundbeziehungen und -begriffe

hinreichend erklären. Auch CARNOT hing anfänglich der Lehre vom Wärmestoff an, revidierte sich jedoch später. Experimente von Sir Benjamin THOMPSON (RUMFORD, 1753–1814) und von Humphry DAVY legten den Schluß nahe, daß es sich bei Wärmeprozessen nicht um Austausch eines Wärmestoffes, sondern um Bewegungs- bzw. Energieumwandlungsvorgänge handle.

Das Prinzip der Erhaltung und Umwandlung der Energie – für mechanische Energie seit langem als gültig vorausgesetzt – wurde von Julius Robert MAYER (1814–1878) ausgesprochen, begründet und durch rechnerische sowie experimentelle Angabe eines Zahlenwertes für das mechanische Wärmeäquivalent belegt. Hermann von HELMHOLTZ (1821–1894) formulierte es umfassender, während James Prescott JOULE (1818–1889), Ludwig August COLDING (1815–1889) u. a. weitere Experimente und Messungen beisteuerten. In der Folgezeit wurden die „Hauptsätze", wichtige Grundlagen der Thermodynamik, formuliert.

Die kinetische Theorie der Gase wurde, anknüpfend u. a. an Überlegungen von Daniel BERNOULLI, von der Mitte des 19. Jh. ab vor allem von Rudolph Julius Emanuel CLAUSIUS (1822–1888) gefördert; er führte auch den Entropiebegriff ein. Die Gasgesetze (Robert BOYLE, 1627–1691; Edmé MARIOTTE, 1620–1684; John DALTON, 1766–1844; Louis-Joseph GAY-LUSSAC, 1778–1850; Amadeo AVOGADRO, 1776–1856) erfuhren ihre statistische Begründung durch Ludwig Eduard BOLTZMANN (1844–1906). Josiah Willard GIBBS (1839–1903) erarbeitete das nach ihm benannte Phasengesetz, James Clerk MAXWELL (1831–1879) beschrieb die Geschwindigkeitsverteilung der Moleküle eines Gases.

1.5. Physik und elektromagnetisches Feld

OERSTEDS Entdeckung (1820), daß elektrischer Strom magnetische Wirkung ausübt, leitete zahlreiche Arbeiten zu den Wechselwirkungen zwischen Elektrizität und Magnetismus ein. Daran waren vor allem an den nach der bürgerlichen Revolution in Frankreich eingerichteten polytechnischen Bildungsstätten tätige Physiker beteiligt.

André Marie AMPÈRE (1775–1836) stellte Regeln für die Beziehungen zwischen Stromrichtung und magnetischer Ablenkung auf, untersuchte Wechselwirkungen zwischen parallelen Strömen, wies nach, daß „natürlicher" und durch Strom hervorgerufener Magnetismus wesensgleich sind und arbeitete um 1825 die Hypothese des durch „Molekularströme" verursachten Ferromagnetismus aus.

Jean Baptiste BIOT (1774–1862) und Félix SAVART (1791–1841) arbeiteten gemeinsam über elektromagnetische Kräfte (Biot-Savartsches Gesetz 1820). Dominique Francois Jean ARAGO (1786–1853) fand den „Rotationsmagnetismus".

Michael FARADAY (1791–1867), ein sehr vielseitiger Physiker und Chemiker, befaßte sich mit der Verflüssigung von Gasen, entdeckte Benzen (Benzol) und Buten (Butylen) sowie grundlegende quantitative Beziehungen für elektrochemische Vorgänge. Seine bedeutendsten Arbeiten galten der Elektrizität und dem Magnetismus. Er untersuchte Zusammenhänge zwischen Polarisationsebene des Lichtes und Magnetismus, die Drehung von Leitern um Magnetpole (und umgekehrt) und Vorgänge im Dielektrikum. Seine für die Elektro- und Energietechnik folgenreichste Leistung war die Entdeckung der elektromagnetischen Induktion (1832).

FARADAY ersetzte die bislang verbreitete Vorstellung einer zwischen Körpern unendlich schnell und ohne Beteiligung des dazwischenliegenden Raumes stattfindenden Fernwirkung durch die Konzeption einer Nahewirkungstheorie. Nach ihr pflanzen sich elektrische und magnetische Wirkungen mit endlicher Geschwindigkeit von Punkt zu Punkt in einem elektrischen bzw. magnetischen Feld fort. FARADAY beschrieb durch elektrische oder magnetische Feldlinien die Felder.

FARADAYS anschaulich-qualitative Vorstellungen waren Ausgang der nach 1860 von MAXWELL entwickelten Theorie des elektromagnetischen Feldes („Elektrizität und Magnetismus", 1873). Sie gibt eine einheitliche, zusammenfassende Darstellung elektrischer, magnetischer und optischer Erscheinungen; aus den Maxwellschen Gleichungen sind die Grundgesetze der Optik, des Magnetismus und der Elektrizität ableitbar. MAXWELL hatte die Existenz elektromagnetischer Wellen mit Frequenzen um Größenordnungen unterhalb derjenigen des sichtbaren Lichtes vorausgesagt. Ihre Entdeckung (1886), Untersuchung und der Nachweis ihrer Wesensgleichheit mit Lichtwellen war das Verdienst von Heinrich HERTZ (1857–1894). Die „Hertzschen Wellen" sind Grundlage der gesamten leiterlosen Informationsübertragung.

Vor allem in der zweiten Hälfte des 19. Jh. wurden Teile des elektromagnetischen Spektrums näher untersucht, wobei die Resultate meist auch weitreichende technische Auswirkungen hatten.

Ausgehend von Beobachtungen und Versuchen Joseph FRAUNHOFERS (1787–1826) wurde als unschätzbares Hilfsmittel wissenschaftlicher Forschung und zahlreicher technologischer Anwendungen die Spektralanalyse vor allem von Robert Wilhelm BUNSEN (1811–1899) und Gustav Robert KIRCHHOFF (1824–1897) geschaffen.

Die Infrarotstrahlung (zunächst der Sonne) wurde von Friedrich Wilhelm HERSCHEL (1738–1822) gefunden, während RITTER entdeckte, daß dem violetten Teil des Lichtspektrums ein ultravioletter Bereich folgt.

Aufsehen erregte die Entdeckung einer „Neuen Art von Strahlen" – so nannte sie Wilhelm Conrad RÖNTGEN (1845–1923) in seiner ersten Veröffentlichung (1895). Innerhalb weniger Wochen nach ihrer Entdeckung hatte RÖNTGEN ihre wichtigsten Eigenschaften festgestellt und schon bald auf Möglichkeiten ihrer medizinischen und technischen Anwendung hingewiesen.

1.6. Neue Erkenntnisse und Resultate

Bereits elektrochemische Vorgänge und Gesetze hatten auf die Existenz einer elektrischen Elementarladung hingewiesen. Sie wurde, ausgehend von Versuchen über den Stromdurchgang durch verdünnte Gase (Julius PLÜCKER, 1801–1868; Johann Wilhelm HITTORF, 1824–1914; William CROOKES, 1832–1919) in der zweiten Hälfte des 19. Jh. gefunden und untersucht.

Die „körnige" Struktur der Katodenstrahlen sowie ihre Ablenkbarkeit durch elektrische und magnetische Felder wurden nachgewiesen (ab 1891 Bezeichnung „Elektronen"); die spezifische Elektronenladung wurde durch Joseph John THOMSON (1856–1940), Wilhelm Carl Werner WIEN (1864–1928) und Emil WIECHERT (1881–1947) bestimmt. 1910 gelang die unmittelbare Messung der Elektronenladung durch Robert Andrews MILLIKAN (1868–1953).

Die natürliche Radioaktivität wurde 1896 an Uraniumsalzen von Henri BECQUEREL (1862–1908) entdeckt. Bei ihrer

Untersuchung fanden Marie CURIE (1867–1934) und Pierre CURIE (1859–1906) die Elemente Polonium und Radium. Aus den Strahlungskomponenten, deren eine Ernest RUTHERFORD (1871–1937) und Frederic SODDY (1878–1956) als Heliumkerne (Alphateilchen) identifizierten, aus „Emanationen" und dem 1899 erkannten Zerfallsgesetz schlossen RUTHERFORD und SODDY auf Zerfall und Umwandlung der Atome als Ursache der natürlichen Radioaktivität. Der experimentelle Nachweis gelang und fand seinen Ausdruck in den radioaktiven Zerfallsreihen.

Bemühungen um eine Formel für die spektrale Energieverteilung erhitzter Körper waren nur für Teilbereiche des Spektrums befriedigend. Max Karl Ernst Ludwig PLANCK (1858–1947) beseitigte diese Schwierigkeiten, indem er die Quantelung der Energie im Strahlungsfeld postulierte. Die von ihm im Jahre 1900 angegebene Strahlungsformel steht in Übereinstimmung mit den Messungen und enthält eine Konstante von der Dimension einer Wirkung. Dieses Wirkungsquantum markierte den Beginn der Quantentheorie und erwies sich in der Folgezeit als von fundamentaler physikalischer Bedeutung.

Albert EINSTEIN (1879–1955) erklärte mit seiner Hilfe den lichtelektrischen Effekt (Wilhelm HALLWACHS, 1859–1922) und entwickelte die Lichtquantentheorie (sie wurde u. a. durch MILLIKAN und 1923 durch Arthur Holly COMPTON, 1892–1962, experimentell bestätigt). Er übertrug seine Vorstellungen auf die Theorie der Wärmekapazität fester Körper und sprach den Gedanken der dialektischen Dualität Welle/Teilchen für elektromagnetische Strahlung aus.

Die Erforschung der Elektronen, Röntgen-Strahlen, der radioaktive Zerfall und die Lichtquantentheorie hatten verdeutlicht, daß mechanische Vorstellungen zur Erklärung des physikalischen Geschehens mit fortschreitendem Erkenntnisstand nicht genügten und insbesondere Atome viel kompliziertere Gebilde waren als ursprünglich angenommen.

Nach wenig befriedigenden Hypothesen mehrerer Physiker stellte RUTHERFORD 1911 sein Planetenmodell mit den um einen Kern kreisenden Elektronen vor. Es erfuhr wesentliche Erweiterungen, indem man es (1913) mit der Quantenhypothese verknüpfte.

Dieses nach Niels Henrik David BOHR (1885–1962) benannte Atommodell konnte u. a. das Linienspektrum des Wasserstoffes erklären. Eine experimentelle Bestätigung Bohrscher Gedankengänge lieferten z. B. die Elektronenstoßversuche von Gustav HERTZ (1887–1979) und James FRANCK (1882–1964).

In der Folgezeit verfeinerten u. a. Arnold SOMMERFELD (1868–1951) und Wolfgang PAULI (1900–1958) das Bohrsche Modell durch Einführung von Nebenquantenzahlen (Ellipsen statt Elektronenkreisbahnen) und der Berücksichtigung relativistischer Massenänderungen (s. u.) bzw. des Ausschlußprinzipes. Durch Werner HEISENBERG (1901–1976), Max BORN (1882–1970) u.a. wurde die Matrizenmechanik, durch Louis de BROGLIE (geb. 1892) und Erwin SCHRÖDINGER (1887–1961) die Wellen- und die Quantenmechanik entwickelt, ausgebaut und auf die Atomtheorie angewandt. Auch das Periodensystem konnte nunmehr physikalisch erklärt werden.

Den Ausgang der speziellen Relativitätstheorie bildete EINSTEINS Arbeit (1905) „Zur Elektrodynamik bewegter Körper". Sie postulierte die Konstanz der Lichtgeschwindigkeit im Vakuum, erübrigte die Annahme eines „ruhenden Äthers", dessen experimenteller Nachweis trotz Aufwandes stets mißlungen war, und stimmte in ihren Schlußfolgerungen mit Beobachtungen und Messungen überein, z. B. Massenzunahme sehr schneller Elektronen (1912).

Die in EINSTEINS Aufsatz „Ist die Trägheit eines Körpers von seinem Energieinhalt abhängig?" vertretenen Thesen wurden durch die weiteren Forschungen nicht nur vielfach bestätigt, sondern erlangten große technische Bedeutung (Kernenergiegewinnung).

Die Gleichheit von träger und schwerer Masse ist eine Grundlage der allgemeinen Relativitätstheorie EINSTEINS (1915). Sie erstreckt das Relativitätsprinzip auf beliebig bewegte Systeme. Von ihr vorhergesagte Effekte konnten durch Messungen bestätigt werden.

Die natürliche Radioaktivität stimulierte die Untersuchung anderer Atomkerne. Bei entsprechenden Experimenten fand RUTHERFORD 1919, daß Stickstoffatome durch Beschießen mit Alphateilchen unter Freisetzung eines Wasserstoffkernes in ein Sauerstoffisotop umgewandelt wurden. Kernumwandlungen konnten also auch künstlich hervorgerufen werden. Dieses Resultat wurde bald an zahlreichen Elementen bestätigt, wobei immer leistungsfähigere Teilchenbeschleuniger gebaut wurden (Kreisbeschleuniger ab 1930, Linearbeschleuniger ab 1931). Die Annahme, daß Atomkerne aus Protonen und Kernelektronen zusammengesetzt seien, wurde 1932 durch die Entdeckung des Neutrons (James CHADWICK, 1891–1974) widerlegt. Sie wurde abgelöst von der Theorie des aus Neutronen und Protonen aufgebauten Kernes (Werner HEISENBERG, Dmitri IWANENKO, geb. 1904). Das von Paul Adrian Maurice DIRAC (1902–1984) vorausgesagte Positron wurde 1932 von Carl David ANDERSON (geb. 1905) in der Höhenstrahlung nachgewiesen. 1937 folgten die von Hideki YUKAWA (1907–1981) zur Deutung des Kernzusammenhaltes vorausgesagten Mesonen.

Irène JOLIOT-CURIE (1897–1956) und Frédéric JOLIOT (1900–1958) entdeckten 1933 bei der Beschießung von Aluminiumfolie die „künstliche" Radioaktivität. Weitere Experimente und Ergebnisse folgten und eröffneten u.a. die Nutzungsmöglichkeiten radioaktiver Isotope. 1937 wurde als erstes künstliches Element „Technetium" hergestellt.

Mit der Bestrahlung schwerer Atomkerne durch Neutronen befaßte sich vor allem Enrico FERMI (1901–1954). Bei der Bestrahlung von Uranium glaubte FERMI Transurane erhalten zu haben. Ähnliche Versuche mit ähnlichen Ergebnissen unternahmen gemeinsam Otto HAHN (1879–1968), Lise MEITNER (1878–1968) und Fritz STRASSMANN (1902–1980). Es stellte sich jedoch heraus, daß die durch Beschluß gewonnenen Stoffe keine Transurane, sondern „Splitter" der durch Neutronenbeschuß gespaltenen Uraniumatome waren.

1939 wiesen Frédéric JOLIOT und Otto HAHN darauf hin, daß bei der Spaltung von Uraniumatomen freiwerdende Neutronen weitere Atome spalten könnten und somit eine energieliefernde Kettenreaktion prinzipiell möglich sein müßte. 1942 wurde sie von FERMI und seinen Mitarbeitern realisiert.

Die damit eröffnete Möglichkeit, Kernenergie zu gewinnen und zu nutzen, wurde 1954 für friedliche Zwecke durch die Inbetriebnahme des ersten Kernkraftwerkes in Obninsk (UdSSR) erschlossen.

Zu den gegenwärtigen Hauptaufgaben der Physik gehören die weitere Erforschung der Atomkerne und der Wechselwirkung zwischen ihren Bestandteilen, die Physik der Elementarteilchen, die Festkörper- und Plasmaphysik und die Erschließung der für die Lösung von Energiefragen besonders wichtigen gesteuerten Kernfusion.

2. Klassische Physik

Unter *klassischer Physik* versteht man diejenigen Gebiete der Physik, die Ende des vorigen Jahrhunderts in ihren Grundlagen im wesentlichen erforscht waren, Mechanik, Elektrizitätslehre, Wärmelehre, Wellenlehre, Optik, Akustik. Zu dieser Zeit war die Vorstellung verbreitet, daß die Grundlagen der Physik vollständig bekannt sind, daß man nur noch quantitative Erweiterungen zu erwarten habe.
Beobachtungen und Überlegungen, die mittels der klassischen Physik nicht erklärt werden konnten bzw. im Widerspruch zu dieser standen, widerlegten aber diese Auffassung. Die Verbreitung und Festigung der Vorstellung des atomaren Aufbaus der Materie (BOLTZMANN), die Entdekkung des Aufbaus der Atome aus Atomkern und Elektronenhülle (RUTHERFORD), die Entwicklung der Vorstellung der Quantelung der Energie (PLANCK) bzw. die nachfolgende Formulierung neuer Bewegungsgleichungen für Prozesse im atomaren Bereich (HEISENBERG, SCHRÖDINGER) und die Entwicklung der Relativitätstheorie mit der Äquivalenz von Masse und Energie als Konsequenz (EINSTEIN) waren der Ausgangspunkt für das Entstehen neuer Gebiete der Physik, die man gern als *moderne Physik* bezeichnet.
Aus dieser Entwicklung darf aber keinesfalls geschlossen werden, daß die klassische Physik überholt oder gar falsch sei. Die Entdeckung allgemeinerer Gesetzmäßigkeiten ist in der Physik mit dem Erkennen neuer Gültigkeitsgrenzen bisher bekannter Gesetze verknüpft. So gilt die Newtonsche Mechanik nur für Geschwindigkeiten, die klein gegen die Vakuumlichtgeschwindigkeit sind. Ebenso gilt sie nur für die Bewegung makroskopischer Körper, nicht für die Bewegung von Elementarteilchen im atomaren Bereich. Die allgemeineren Gesetzmäßigkeiten schließen die bisher bekannten Gesetze mit ein. So vereinfacht sich die Bewegungsgleichung der relativistischen Mechanik für kleine Geschwindigkeiten zur Newtonschen Bewegungsgleichung. Ebenso gewinnt man für makroskopische Objekte aus der Schrödinger-Gleichung die Newtonsche Bewegungsgleichung zurück. Das heißt, die klassische Physik ist weiterhin gültig, es gibt nur Problemstellungen, auf die sie nicht anwendbar ist. Mehr noch, der überwiegende Anteil der technischen Anwendungen der Physik bewegt sich im Gültigkeitsbereich der klassischen Physik.
In dem Teil klassische Physik wird Wert darauf gelegt, die Grundbegriffe genau zu erläutern. Es ist ein Teil der physikalischen Kenntnis, daß man jeden Begriff mit Beobachtungen und Vorstellungen koppelt und nicht wie eine Vokabel lernt. Die Grundgesetze werden von experimentellen Erfahrungen ausgehend formuliert. Aus der Mannigfaltigkeit der Lösungen werden typische Beispiele herausgegriffen. Die Darstellung ist möglichst elementar gewählt, versucht aber auch den Zusammenhang zum Allgemeinen herzustellen. Die physikalischen Gesetze sind mathematische Gleichungen. Daher muß einer Vertiefung der physikalischen Kenntnisse eine Erweiterung der mathematischen Fähigkeiten parallellaufen. Dort, wo komplizierte Beispiele besprochen werden, wird der Leser in der Regel auch auf anspruchsvollere Mathematik stoßen. Es ist nicht sinnvoll, die Physik in elementare und höhere Physik aufzuteilen. Abschnitte mit elementaren und anspruchsvolleren Betrachtungen wechseln sich ab. Teilweise wird von der mathematischen Durchdringung abgesehen bzw. wird nur die Formel angegeben und der physikalische Inhalt allein verbal erläutert.

2.1. Mechanik

Die Aufgabe der Mechanik besteht in der Erklärung von Bewegungsabläufen von Massen aus dem diese Bewegungen bestimmenden Naturgesetz heraus. In der Physik werden solche Abläufe mathematisch beschrieben. Dazu wird die Abhängigkeit der Position von der Zeit, die *Bahnkurve*, durch eine Funktion oder ein Diagramm angegeben. Es muß dabei festgelegt werden, wieviel solche Funktionen der Zeit notwendig sind, um den Ablauf vollständig zu beschreiben. Durch diese Überlegung wird die Zahl der *Freiheitsgrade* des Systems ermittelt. Ist diese Funktion der Zeit bekannt, dann kann daraus abgeleitet werden, welche Geschwindigkeit ein Objekt hat oder wie es seine Geschwindigkeit ändert. Diese Zusammenhänge werden in der *Kinematik* untersucht.
In der Mechanik wird untersucht, warum ein bestimmter Bewegungsablauf auftritt, welches Naturgesetz ihn festlegt. Es interessiert die *Dynamik* der Bewegung. Um dieses Gesetz formulieren zu können, muß man Experimente auswerten; die physikalischen Gesetze sind Erfahrungsgesetze. Das Ergebnis ist das Newtonsche Gesetz, die Grundgleichung der Mechanik. Es besagt, daß die Kräfte die Impulsänderung einer Masse bestimmen.
Es gibt sehr viele experimentelle Erfahrungen. Die theoretische Durchdringung eines Gebietes ist um so weiter vorangeschritten, je besser es gelungen ist, diese Erfahrungen als Lösungen weniger Grundgleichungen oder eines Grundgesetzes zu verstehen. In diesem Sinne ist die Mechanik weit entwickelt. Alle Erscheinungen können von der Newtonschen Gleichung ausgehend verstanden werden; das gilt für das Fallgesetz, das Hebelgesetz und den Energiesatz der Mechanik genauso wie für die Schwingungsgleichung einer Saite oder die Bernoullische Gleichung über das Strömen einer Flüssigkeit. Das heißt nicht, daß man bei der Lösung jedes konkreten Problems von der Newtonschen Gleichung ausgeht. Man benutzt vielfach für Teilgebiete daraus abgeleitete Gleichungen, die oft zunächst als Grundgleichung dieses Teilgebietes experimentell ermittelt worden sind. Wenn das physikalische Verständnis entwickelt ist, wird man aber diese Einzelaussagen nicht wie Vokabeln auswendig lernen müssen, sondern als Modifikationen oder Lösungen der Grundgleichung verstehen.
Es gibt in der Natur Gesetze, die in mehreren oder allen Gebieten der Physik gültig sind, nicht nur in der Mechanik. Diese Gesetze müssen daher als Aussagen der Newtonschen Gleichung für mechanische Systeme auffindbar sein. Es ist deshalb notwendig, die Grundgleichung der Mechanik so umzuformen, daß man von den Besonderheiten eines mechanischen Systems möglichst abstrahieren kann,

um Gleichungen zu gewinnen, aus denen man Schlußfolgerungen über die Mechanik hinaus ziehen kann. Die Theoretische Mechanik, als Theorie eines besonders gut bekannten Gebietes der Physik, übernimmt die Rolle einer *Modelltheorie*. Diese Fragen werden in dem Kapitel „Verallgemeinerte Formulierungen" behandelt.

2.1.1. Grundgrößen

Physikalische Zusammenhänge, wie zwischen Position und Zeit, müssen zahlenmäßig festgehalten werden, um die zugrunde liegenden Gesetze erkennen bzw. um gemessene und berechnete Werte vergleichen zu können. Man muß dabei sehen, daß es physikalische Größen mit verschiedener *Dimension* gibt, wie Längen, Zeiten. Vergleichen kann man nur Größen gleicher Dimension, Längen mit Längen, nicht aber verschiedener Dimension, also nicht Längen mit Zeiten. Um eine Länge zu messen, benötigt man einen Maßstab zum Vergleichen, es muß eine Einheit festgelegt worden sein.

Eine *Einheit* kann willkürlich gewählt werden. Es ist aber unzweckmäßig, für alle physikalischen Größen, die sich in ihrer Dimension unterscheiden, eine willkürliche Einheit zu definieren. Wenn man für Längen das Meter (m) und für Zeiten die Sekunde (s) als Einheit definiert hat, dann ist es am übersichtlichsten, für die Geschwindigkeit, die die Dimension Länge/Zeit hat, die Einheit m/s zu wählen. Man spricht dann von abgeleiteten Einheiten. Es gibt also physikalische Größen, deren Einheit definiert wird – in der Mechanik müssen das drei sein – und physikalische Größen mit abgeleiteten Einheiten. Welche (drei) Basisgrößen man wählt, ist in gewissem Maße willkürlich. Je nach der Wahl unterscheidet man verschiedene Maßsysteme. Das gesetzlich verbindliche Internationale Einheitensystem (SI) geht von den Basisgrößen Länge, Zeit und Masse aus und verwendet die Einheiten Meter, Sekunde und Kilogramm. Abgeleitete Einheiten sind also Kombinationen von Meter, Sekunde und Kilogramm.

Man strebt an, daß Einheiten international einheitlich gewählt werden. Das erleichtert den Handel wie auch die wissenschaftliche Diskussion. Umrechnungen, wie sie jetzt noch zwischen Meter und yard oder Hektoliter und barrel beim Handel mit einigen Ländern erforderlich sind, bedeuten nur Erschwernis. Weiterhin verlangt man von einer Einheit, daß sie reproduzierbar ist, daß sie nicht verloren gehen kann. Sie soll möglichst leicht zugänglich sein, damit durch Vergleich mit der definierten Einheit und nicht durch Vergleich mit Kopien geeicht werden kann, wobei sich letzteres natürlich nur im beschränkten Maße durchsetzen läßt. Schließlich sollen die Einheiten so festgelegt werden, daß eine physikalische Größe mit dem Zahlenwert Eins eine aus dem täglichen Leben vertraute Größenordnung hat. Um sehr große und sehr kleine Zahlenwerte zu vermeiden, hat man daher Vorsätze zu den Einheiten festgelegt, die Abkürzungen für bestimmte negative und positive Zehnerpotenzen sind (Tab. 2.1.1–1).

Tab. 2.1.1–1 Abkürzungen für Zehnerpotenzen

Bezeichnung	Zeichen	Bedeutung
Atto	a	10^{-18}
Femto	f	10^{-15}
Piko	p	10^{-12}
Nano	n	10^{-9}
Mikro	µ	10^{-6}
Milli	m	10^{-3}
Zenti	c	10^{-2}
Dezi	d	10^{-1}
Deka	da	10^{1}
Hekto	h	10^{2}
Kilo	k	10^{3}
Mega	M	10^{6}
Giga	G	10^{9}
Tera	T	10^{12}
Peta	P	10^{15}
Exa	E	10^{18}

Eine physikalische Größe besteht also aus einem *Zahlenwert* und einer *Einheit*. Beides steckt zusammengefaßt im *Symbol* für diese physikalische Größe. Wird z. B. eine Strecke mit l bezeichnet, und hat diese Strecke eine Länge von fünf Meter, dann schreibt man $l = 5$ m. Weitere Beispiele sind $t = 10{,}21$ s und $m = 520$ kg. Aus der zweiten Gleichung entnimmt man, daß t eine Größe mit der Dimension Zeit ist. Gemessen in der Einheit Sekunde ist ihr Zahlenwert 10,21. Die Symbole physikalischer Größen werden immer kursiv gedruckt, so daß keine Verwechslungen mit Einheiten auftreten können. In allen Gleichungen sind bei der zahlenmäßigen Auswertung für physikalische Größen immer Zahlenwert und Einheit einzusetzen.

In anderen Gebieten der Physik kommen zu den drei in der Mechanik erforderlichen Einheiten weitere hinzu, die Temperatur (Kelvin), die Menge (Mol), die elektrische Stromstärke (Ampere) und die Lichtstärke (Candela).

2.1.1.1. Länge, Fläche, Volumen, Winkel. Länge. Die Einheit der *Länge* ist das Meter (m). Um seine Reproduzierbarkeit zu sichern, war es zunächst an den Ausmaßen der Erde definiert worden. Es sollte der zehnmillionste Teil des Viertelmeridians Nordpol–Äquator, der durch die Pariser Sternwarte verläuft, sein. Nach dieser Definition wurde ein Urmeter geschaffen, ein Metallstab aus Platin (90 %) und Iridium (10 %). Das Meter war der Abstand zwischen zwei Markierungen auf diesem Stab bei bestimmten Bedingungen von Druck und Temperatur.

Es ist verständlich, daß man bei Verbesserung der Meßtechnik feststellte, daß der Viertelmeridian nicht genau 10 000 000 m lang ist. Da es nicht sinnvoll ist, bei jeder Verbesserung der Meßtechnik das Meter neu festzulegen, mußte man sich damit abfinden, daß man keine genau reproduzierbare Längeneinheit eingeführt hatte. Das Meter war allein durch das Urmeter verkörpert. Von diesem wurden Kopien an die damals an dem Meterabkommen beteiligten 31 Staaten verteilt. Von diesen ausgehend wurden sogenannte Endmaße verschiedener Länge hergestellt, die in ihrer Genauigkeit besser als 10^{-3} mm waren.

Um andere Größenordnungen zu charakterisieren, verwendet man die in Tab. 2.1.1–1 zusammengestellten Vorsätze.

Um eine genauer reproduzierbare Einheit zu schaffen und um eine im Labor direkt *zugängliche Einheit* zu haben, galt von 1960 bis 1983 die Festlegung: Ein Meter ist das 1 650 763,73fache der Wellenlänge, die das zum Übergang $5d_5$-$2p_{10}$ des ^{86}Kr-Atoms gehörige Licht im Vakuum besitzt.

1983 wurde festgelegt, der Vakuumlichtgeschwindigkeit den festen Wert $c = 299\,792\,458$ m/s zuzuordnen. Damit lautet die Definition des Meters: Das Meter ist die Länge der Strecke, die Licht im Vakuum während der Dauer von

1/299 792 458 Sekunden durchläuft. Wesentlich für diese Neudefinition war die Möglichkeit, Laser hoher Frequenzstabilität – in der Genauigkeit dem Caesium-Standard (vgl. 2.1.1.2.) vergleichbar – herzustellen.
In der physikalischen Forschung verwendet man Vergleichsmaße, die aus Kombinationen von Naturkonstanten gebildet werden und die die physikalische Anschauung fördern. So charakterisiert der Bohrsche Wasserstoffradius $a_0 = \hbar^2/(m e^2)$ (vgl. 4.2.2.) die Ausdehnung eines Atoms. Die Parallaxensekunde ist die Entfernung, die ein Stern hat, der über einer dem Erdbahnradius entsprechenden Strecke unter einem Winkel von einer Sekunde erscheint.
Flächen sind, geometrisch gesehen, zweidimensionale Gebilde. Sie haben z. B. Länge und Breite. Die physikalische Größe Flächeninhalt hat die Dimension (Länge)^2. Die abgeleitete Einheit ist daher Quadratmeter (Tab. 2.1.1–2).

Tab. 2.1.1–2 Flächeneinheiten (Auswahl)

Bezeichnung	Zeichen	Bedeutung
Quadratmillimeter	mm^2	10^{-6} m^2
Quadratzentimeter	cm^2	10^{-4} m^2
Quadratmeter	m^2	
Hektar	ha	10^4 m^2
Quadratkilometer	km^2	10^6 m^2

Volumina sind, geometrisch gesehen, dreidimensionale Gebilde. Sie haben z. B. Länge, Breite und Höhe. Die physikalische Größe Rauminhalt (Volumen) hat die Dimension (Länge)^3. Die abgeleitete Einheit ist daher Kubikmeter (Tab. 2.1.1-3).

Tab. 2.1.1–3 Volumeneinheiten (Auswahl)

Bezeichnung	Zeichen	Bedeutung
Kubikmillimeter	mm^3	10^{-9} m^3
Kubikzentimeter	cm^3	10^{-6} m^3
Kubikdezimeter	dm^3	10^{-3} m^3
Liter	l	10^{-3} m^3
Hektoliter	hl	10^{-1} m^3
Kubikmeter	m^3	
Kubikkilometer	km^3	10^9 m^3

Winkel. Eine direkte *Längenmessung* erfordert das Anlegen eines Maßstabes an die zu messende Strecke. Das ist nicht immer möglich. Oft werden indirekte Messungen (Abb. 2.1.1–4) durchgeführt, z. B. Winkel gemessen. So kann die Entfernung AB durch Messung der Strecke AA' und die

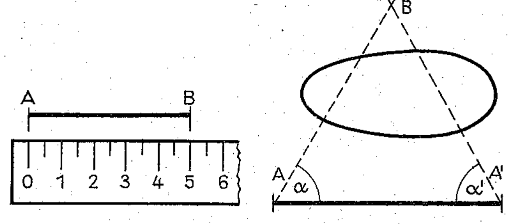


Abb. 2.1.1–4 Direkte und indirekte Messung

Winkel α und α' ermittelt werden. Es werden dabei aber immer Voraussetzungen hineingesteckt, deren Richtigkeit nicht generell feststeht. So wird z. B. angenommen, daß die Lichtstrahlen mit den Geraden der euklidischen Geometrie identisch sind.
Der ebene Winkel wird als Kreisbogenlänge dividiert durch den Radius definiert (Abb. 2.1.1–5). Der Winkel ist eine dimensionslose Größe. Als ergänzende Einheit, d. h. um zu kennzeichnen, daß man von einem Winkel spricht, kann man *Radiant* (rad) hinzufügen. Auch das *Gradmaß* wird verwendet. Die Umrechnung ist aus Tab. 2.1.1–6 zu ersehen. Bei dem in der Geodäsie benutzten Neugrad wird der volle Winkel in 400 Gon unterteilt.

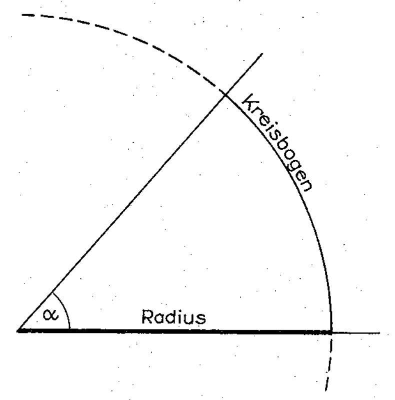


Abb. 2.1.1–5 Definition des Winkels

In der Physik hat man oft Raumwinkel zu betrachten:

$$\text{Raumwinkel} = \frac{\text{Kugelflächenausschnitt}}{(\text{Radius})^2}.$$

Da die Kugeloberfläche den Flächeninhalt $4\pi r^2$ besitzt, hat ein voller Raumwinkel den Wert 4π, ein Halbraum den Raumwinkel 2π. Ein Raumwinkel kann durch *Steradiant* (sr) gekennzeichnet werden.
2.1.1.2. Zeit. Physikalische Vorgänge, unser Leben, die Entwicklung der Gesellschaft, laufen in der Zeit ab. In diesem Sinne ist uns der Begriff Zeit vertraut. Um in der Physik damit arbeiten zu können, braucht man eine Meßvorschrift und eine Einheit. Zum Messen einer Zeitdauer vergleicht man diese mit periodischen Vorgängen, für die man voraussetzt, daß jede Periode die gleiche Zeit benötigt. Es war naheliegend, den täglichen Rhythmus als Maßstab zu nehmen. Als *Zeiteinheit* wurde der Sterntag gewählt, also die Zeit, die die Erde für eine Umdrehung gegenüber dem Fixsternhimmel benötigt. Der wahre Sonnentag, d. h. die Zeit zwischen zwei aufeinanderfolgenden Kulminationen der Sonne, hat dagegen zu verschiedenen Zeiten des Jahres

Tab. 2.1.1–6 Winkelmaße

Bezeichnung	Bogenmaß	Gradmaß
Radiant	1 rad	57,29578° = 57°17′45″
Grad	0,017453 rad	1°
voller Winkel	2π	360°
rechter Winkel	$\pi/2$	90°

eine unterschiedliche Dauer. Daher definierte man, vom Sterntag ausgehend, einen mittleren Sonnentag, der in 24 Stunden zu je 60 Minuten zu je 60 Sekunden unterteilt wurde.
Man weiß heute aber, daß die Drehung der Erde auch gewissen Schwankungen unterworfen ist. Man griff daher, wie bei der Definition der Längeneinheit, auf atomare Prozesse zurück und glaubt, so eine stets reproduzierbare Einheit definiert zu haben (Tab. 2.1.1–7):

Tab. 2.1.1–7 Häufig auftretende Zeiteinheiten

Bezeichnung	Zeichen	Bedeutung
Pikosekunde	ps	10^{-12} s
Nanosekunde	ns	10^{-9} s
Mikrosekunde	µs	10^{-6} s
Millisekunde	ms	10^{-3} s
Sekunde	s	
Minute	min	60 s
Stunde	h	3 600 s
Tag	d	86 400 s

9 192 631 770 Perioden der Strahlung, die dem Übergang zwischen den beiden Niveaus der Hyperfeinstruktur des Grundzustandes des Caesiumatoms 133 entspricht, dauern 1 Sekunde.
Nach der Festlegung der Zeiteinheit kann man die Dauer eines Vorganges messen, man kann auch die Dauer verschiedener Vorgänge miteinander vergleichen. Es ist aber auch von Interesse, zu bestimmen, ob ein Ereignis früher bzw. später als ein anderes oder *gleichzeitig* mit einem anderen stattgefunden hat. Geschehen beide Vorgänge am gleichen Ort, so kann man dies problemlos feststellen. Will man jedoch Ereignisse an verschiedenen Orten miteinander vergleichen, sind einige Vorkehrungen nötig.
Man kann an einem Ort zwei Uhren eichen und ihren Gang vergleichen. Dann transportiert man die eine Uhr an den anderen Ort. Oder man eicht die beiden Uhren an verschiedenen Orten, indem man in jedem Labor auf die Definition der Sekunde zurückgreift, dann muß man aber mittels eines Signals zwischen diesen beiden Orten die Uhren gleichstellen. Im ersten Fall treten nur dann keine Schwierigkeiten auf, wenn der Gang der Uhr durch die Bewegung (Beschleunigen, Transport mit konstanter Geschwindigkeit, Abbremsen) nicht beeinflußt wird. Im zweiten Fall treten keine Probleme auf, wenn man ein Signal zur Verfügung hat, das sich unendlich schnell ausbreitet, so daß zwischen Aussenden und Empfang keine Zeit verstrichen ist. Die klassische (nichtrelativistische) Mechanik geht davon aus, daß die genannten Probleme nicht auftreten, daß das Gleichstellen zweier Uhren an verschiedenen Orten problemlos ist. In der klassischen Mechanik gibt es demzufolge eine *absolute Zeit*, die an allen Orten und in allen Bezugssystemen gleich abläuft. Beobachtungen an Objekten, die sich mit Geschwindigkeiten nahe der Lichtgeschwindigkeit bewegen, haben aber gezeigt, daß dieser Standpunkt korrigiert werden muß. Dies geschieht in der speziellen Relativitätstheorie (vgl. 3.1.).

2.1.1.3. Masse. Masse ist die Eigenschaft eines Körpers, einer Beschleunigung einen Widerstand entgegenzusetzen (*träge Masse*), wie das im Newtonschen Grundgesetz zum Ausdruck kommt. Andererseits bestimmt die Masse die Stärke der Gravitationskraft zwischen zwei Körpern (*schwere Masse*). Die Erfahrung lehrt, daß beide Massen gleich sind. Als denkbarer Proportionalitätsfaktor kann Eins gesetzt werden. Dies äußert sich z. B. darin, daß alle Körper gleich schnell fallen, wenn man den Luftwiderstand ausschaltet. Es hat lange gedauert, bis man sich über diese Gleichheit gewundert hat, obwohl von der elektrischen Kraft her, der Coulomb-Kraft, ein anderes Verhalten bekannt ist. Man kann ohne weiteres Körper realisieren, die die gleiche (träge) Masse haben, aber verschieden geladen sind. Die Diskussion dieses Sachverhaltes führte zur allgemeinen Relativitätstheorie.
Masse ist – neben Länge und Zeit – die dritte Grundgröße in der Mechanik. Die gesetzlich vorgeschriebene Einheit ist das Kilogramm (kg):
1 kg ist die Masse des Platin(90 %)-Iridium(10 %)-Zylinders, der in Paris aufbewahrt wird.
Die Masse dieses Zylinders sollte ursprünglich gleich der Masse eines Kubikdezimeters Wasser bei 4 °C sein. Spätere genauere Messungen zeigten, daß diese Wassermenge eine etwas geringere Masse hat.
Die Masseneinheit ist also noch nicht auf atomare Eigenschaften zurückgeführt worden. Da man in der Massenspektroskopie jedoch Verhältnisse von atomaren Massen genauer bestimmen kann als die Absolutwerte, hat man schon seit langem *relative Atommassen* eingeführt. Bei der jetzt verwendeten Vergleichsbasis wird die Masse eines Kohlenstoffatoms ^{12}C gleich zwölf atomaren Masseneinheiten (u) gesetzt (Tab. 2.1.1–8).

Tab. 2.1.1–8 Masse einiger Objekte

Objekt	Masse
Elektron	$9{,}10953 \cdot 10^{-31}$ kg
Proton	$1{,}67265 \cdot 10^{-27}$ kg
Neutron	$1{,}67495 \cdot 10^{-27}$ kg
Atom des Kohlenstoffisotops ^{12}C	$12 \cdot 1{,}66057 \cdot 10^{-27}$ kg
Mond	$7{,}35 \cdot 10^{22}$ kg
Erde	$5{,}975 \cdot 10^{24}$ kg
Sonne	$1{,}99 \cdot 10^{30}$ kg

Von der Masse ist die *Menge* (vgl. 2.6.1.) zu unterscheiden, deren Einheit das Mol (mol) ist.
Dichte ist eine abgeleitete Größe (Dichte = Masse/Volumen). Die Einheit der Dichte ist kg/m³. Oft verwendet man g/cm³, da dann der Zahlenwert überschaubarer ist.

2.1.2. Kinematik

2.1.2.1. Freiheitsgrade. Will man die Bewegung einer Punktmasse, eines Systems von Punktmassen oder eines Körpers beschreiben, so muß man die Lage genau charakterisieren. Man muß sich dazu zuerst überlegen, wieviel Zahlenangaben man benötigt, um die Lage eindeutig zu beschreiben. Diese Anzahl nennt man die Zahl der *Freiheitsgrade f* eines Systems.
Punktmasse. Die Punktmasse ist eine Abstraktion von einer Massenverteilung, bei der die Gesamtmasse als im Schwerpunkt vereinigt gedacht wird und zur Lagebestimmung nur die Freiheitsgrade des Schwerpunktes betrachtet werden. Von allen übrigen Freiheitsgraden des Systems wird abgesehen. Bei der Berechnung der Planetenbahnen um die Sonne, um ein Beispiel zu nennen, ersetzt man die

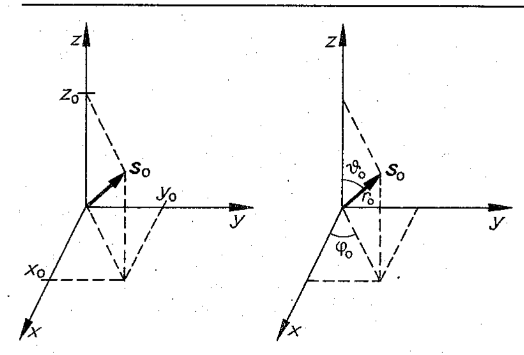


Abb. 2.1.2–1 Koordinaten eines Punktes s_0

Planeten durch Punktmassen. Natürlich kann man in diesem Bild nicht Ebbe und Flut diskutieren. Eine Punktmasse, die sich im Raum frei bewegen kann, hat drei Freiheitsgrade (Abb. 2.1.2–1). Die Lage ist durch den Ortsvektor s_0 bestimmt, der – je nach Wahl der Koordinaten – z.B. die drei Zahlenangaben x_0, y_0 und z_0 bzw. r_0, ϑ_0 und φ_0 umfaßt. Ist der Punkt an eine Fläche gebunden, so hat er nur zwei Freiheitsgrade. Die Position eines Schiffes wird durch Längen- und Breitengrad angegeben. Ein System von zwei Punktmassen, die sich frei im Raum bewegen können, hat sechs Freiheitsgrade, man benötigt zwei Ortsvektoren. Wird der Abstand zwischen diesen beiden Massen festgehalten, dann vermindert sich die Zahl der Freiheitsgrade auf fünf.

Starre Körper haben sechs Freiheitsgrade (vgl. 2.1.11.). Man sieht bei ihnen von der Verschiebung der Massenelemente gegeneinander durch Deformation des Körpers ab.

Mit der Bestimmung der Zahl der Freiheitsgrade eines Systems ist oft ein beträchtliches Maß an Abstraktionsvermögen verknüpft. Zur Beschreibung des Sachverhaltes reicht oft die Angabe weniger Lagekoordinaten aus, obwohl das System viel mehr Freiheitsgrade hat. So interessiert bei der Angabe der Position eines Schiffes z. B. nicht, in welchem Raum sich die Passagiere gerade aufhalten. Für die Fragestellung sind also die „inneren Freiheitsgrade" des Systems Schiff unwesentlich. Bei einem Rennen ist nur die Reihenfolge der Läufer von Bedeutung und wessen Brust als erste die Ziellinie überquert. Diese Angaben machen daher auch im wesentlichen den Inhalt einer Reportage aus. Es ist weniger interessant, wie der Läufer gerade seine Beine, Arme, Finger usw. bewegt.

Schwerpunkt. Der Schwerpunkt ist ein ausgezeichneter Punkt einer Massenverteilung, dessen Bewegung einer einfachen Bewegungsgleichung genügt (vgl. 2.1.3.1.). Er wird daher häufig zur Angabe der Position eines Körpers herangezogen. Für ein System von N Punktmassen, die sich an Orten $\boldsymbol{r}_i$ befinden, ist er durch

$$s = \frac{1}{M}\sum_{i=1}^{N} m_i \boldsymbol{r}_i, \quad M = \sum_{i=1}^{N} m_i$$

gegeben. Für die beiden Massen in Abb. 2.1.2–2 ist es der Punkt, an dem man den Stab unterstützen muß, damit die Wirkung der Schwerkraft ausgeglichen wird und das System im Gleichgewicht ist. Für eine kontinuierliche Massenverteilung, charakterisiert durch die Massendichte $\varrho(\boldsymbol{r})$, ist der Schwerpunkt durch $s = \frac{1}{M}\int \mathrm{d}\tau\varrho(\boldsymbol{r})\,\boldsymbol{r}$, $M = \int \mathrm{d}\tau\varrho(\boldsymbol{r})$ gegeben. Statt der Summe über die Einzelmassen, tritt ein Integral über die Volumenelemente $\mathrm{d}\tau$, in denen sich die Massen $\mathrm{d}\tau\varrho(\boldsymbol{r})$ befinden, auf.

2.1.2.2. Bahnkurve. Um den Bewegungsablauf eines mechanischen Systems vollständig zu kennen, muß in dem interessierenden Zeitbereich die Position bekannt sein. Bei f Freiheitsgraden des Systems erfordert das die Angabe von f Funktionen der Zeit. Diese Gesamtheit wird als Bahnkurve bezeichnet. Für eine einzelne Punktmasse im Raum ist dies der *Ortsvektor* als Funktion der Zeit

$$s(t) = x(t)\boldsymbol{e}_x + y(t)\boldsymbol{e}_y + z(t)\boldsymbol{e}_z,$$

der in kartesischen Koordinaten die drei Zeitfunktionen $x(t)$, $y(t)$ und $z(t)$ umfaßt. Als Teilinformation steckt in der Bahnkurve die *geometrische Kurvenform*, z.B. Gerade, Kreis, Parabel, ohne Angabe des Zeitablaufes. Man erhält die Kurvenform, indem man die Zeit eliminiert, z. B. $z = z(t)$ nach $t = t(z)$ auflöst, und diesen Zusammenhang in $x(t)$ und $y(t)$ für t einsetzt. Die Beschreibung und Diskussion einer Bahnkurve kann man durch *geeignete Wahl des Bezugssystems* entscheidend vereinfachen.

Bewegung auf einer Geraden nach $s(t) = s_0 + \lambda(t)\boldsymbol{e}$ (Abb. 2.1.2–3a). Ist $s_0 = s(t_0)$, dann gilt $\lambda(t_0) = 0$. Legt man den Koordinatenursprung in s_0 und die x-Achse in Richtung $\boldsymbol{e}$, dann vereinfacht sich diese Bahnkurve zu $s(t) = x(t)\boldsymbol{e}_x$ mit $x(t_0) = 0$. Statt eine Gerade im Raum betrachten zu müssen, reicht also die Diskussion der eindimensionalen Bewegung $x(t)$ aus.

Bewegung auf einem Kreis. Hierbei legt man den Koordinatenursprung in den Kreismittelpunkt und zwei Achsen, z. B. x- und y-Achse, in die Kreisebene (Abb. 2.1.2–3b). Die Position auf dem Kreis ist durch Angabe des Winkels $\varphi(t)$ bestimmt. Daher enthält die Bahnkurve diese eine Zeitfunktion

$$s(t) = \boldsymbol{e}_x R\cos\varphi(t) + \boldsymbol{e}_y R\,\sin\varphi(t) \text{ bzw.}$$
$$x(t) = R\cos\varphi(t), \quad y(t) = R\sin\varphi(t).$$

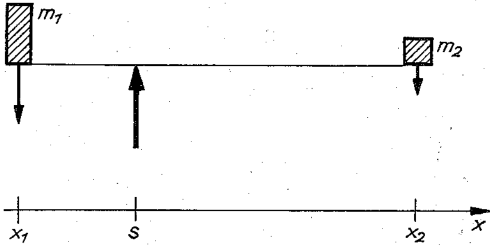


Abb. 2.1.2–2 Schwerpunkt zweier Massen

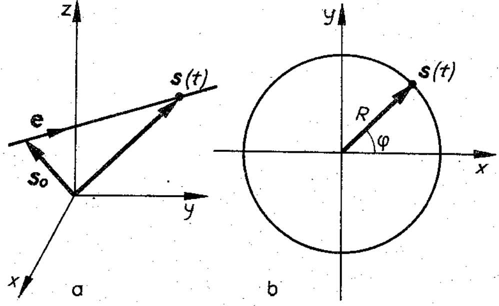


Abb. 2.1.2–3 Bewegung auf (a) einer Geraden und (b) einem Kreis

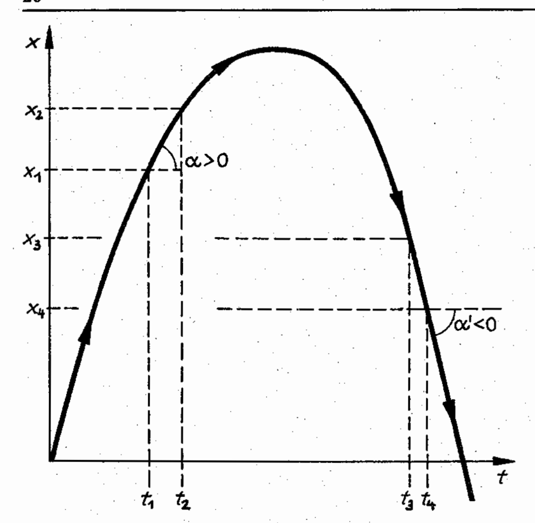


Abb. 2.1.2–4 Bahnkurve $x(t)$ für eine Bewegung auf der x-Achse

2.1.2.3. Geschwindigkeit, Beschleunigung. Geschwindigkeit. Abb. 2.1.2–4 zeigt eine Bahnkurve $x(t)$ für eine Bewegung auf der x-Achse. Aus der Änderung der Position $\Delta x = x(t_2) - x(t_1) = x_2 - x_1$ im Zeitintervall $\Delta t = t_2 - t_1$ kann man die Geschwindigkeit $v = \Delta x/\Delta t$ bestimmen. Man sieht aus dieser Definition, daß die Geschwindigkeit die Dimension Länge/Zeit hat. Gebräuchliche Einheiten sind m/s und km/h, 1 km/h = 1 000 m/3 600 s = 0,278 m/s. Ein Sprinter, der 100 m in 10 s durchläuft, hat eine Durchschnittsgeschwindigkeit von 36 km/h.

Die Messung der Geschwindigkeit erfolgt für ein endliches Zeitintervall Δt, in ihm wird die Durchschnittsgeschwindigkeit ermittelt. In der mathematischen Formulierung wird aus der Bahnkurve $x(t)$ durch den Grenzübergang $\Delta t \to 0$ die Geschwindigkeit zu jedem Zeitpunkt berechnet. Man schreibt $v(t) = dx/dt \equiv \dot{x}(t)$. Aus Abb. 2.1.2–4 ist ersichtlich, daß die Geschwindigkeit proportional zum Tangens des Winkels α ist. Im Grenzübergang ist das der Winkel der Tangente im Punkt $x(t)$ zur t-Achse. Man kann also aus der Abbildung einen qualitativen Eindruck von der Geschwindigkeit erhalten. Je steiler die Kurve ist, desto größer ist die Geschwindigkeit. Im Scheitelpunkt verschwindet die Geschwindigkeit, mit $\alpha = 0$ wird auch $\tan \alpha = 0$. Auf dem rechten Ast der Bahnkurve in Abb. 2.1.2–4 ist die Geschwindigkeit negativ. Mit wachsender Zeit verändert sich die Position nach kleineren x-Werten hin. Der Tangens des Winkels α' ist negativ.

Beschleunigung. Die Beschleunigung beschreibt die Änderung der Geschwindigkeit mit der Zeit. Sie ist über $a = \Delta v/\Delta t$ definiert. Ihre Dimension ist also Länge/(Zeit)^2, ihre Einheit ms^{-2}. Im Grenzübergang $\Delta t \to 0$ kann sie aus der Bahnkurve durch zweimalige Differentiation nach der Zeit gewonnen werden

$a = dv/dt = d^2x/dt^2 \equiv \ddot{x}(t)$.

2.1.2.4. Weg-Zeit-Diagramm. Darstellungen vom Typ der Abb. 2.1.2–4 nennt man Weg-Zeit-Diagramm. Sie ermöglichen einen guten qualitativen Überblick über die Bewegung, und man kann auch quantitative Aussagen entnehmen. Für einige einfache aber häufig wiederkehrende

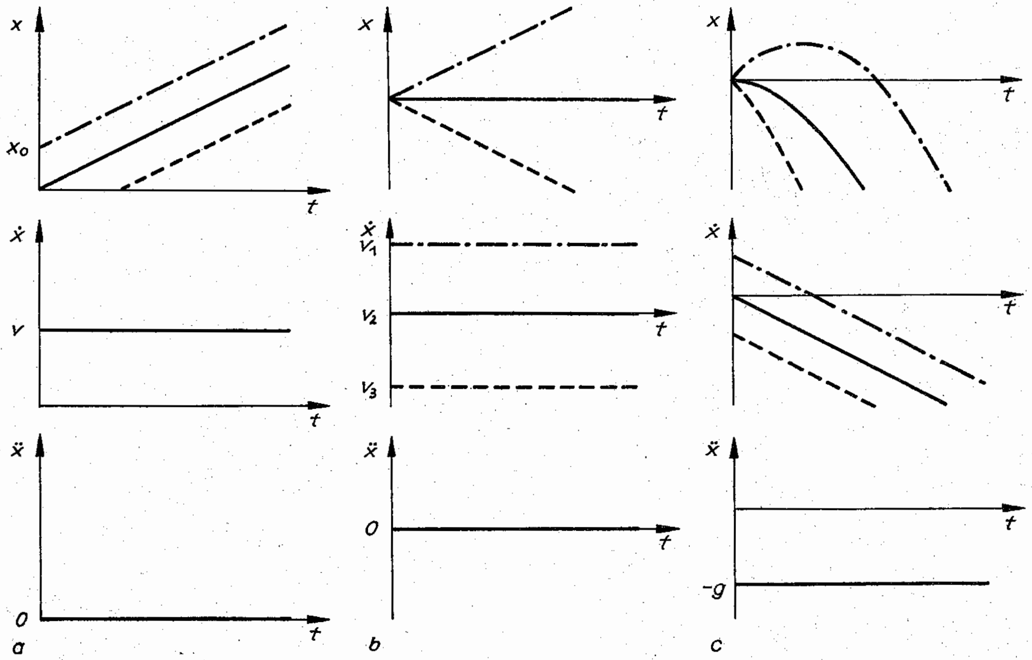


Abb. 2.1.2–5 Weg-Zeit-, Geschwindigkeits-Zeit- und Beschleunigungs-Zeit-Diagramm für verschiedene Anfangsbedingungen und Beschleunigungen; (a) unterschiedliches x_0, gleiches v_0, $a = 0$, (b) gleiches x_0, unterschiedliches v_0, $a = 0$, (c) gleiches x_0, unterschiedliches v_0, $a = -g$.

Beispiele sollen Ort, Geschwindigkeit und Beschleunigung diskutiert werden.
Die Bahnkurve für eine geradlinige Bewegung mit konstanter Geschwindigkeit lautet
$x(t) = vt + x_0$, $\dot{x}(t) = v$, $\ddot{x}(t) = 0$.
Abb. 2.1.2–5a zeigt Bahnkurven gleicher Geschwindigkeit, aber unterschiedlicher Anfangsposition x_0. Die konstante Geschwindigkeit äußert sich im Weg-Zeit-Diagramm dadurch, daß die Bahnkurve eine Gerade ist. In jedem Zeitpunkt ist die Tangente an die Bahnkurve identisch mit der Geraden selbst, der Winkel mit der t-Achse ist zu allen Zeiten gleich dem Steigungswinkel der Geraden. Zu parallelen Geraden gehören gleiche Geschwindigkeiten, nur der Anfangsort $x_0 = x(0)$ ist verschieden. Da die Geschwindigkeit sich nicht ändert, ist die Beschleunigung Null.
In Abb. 2.1.2–5b werden Bahnkurven verschiedener Geschwindigkeiten miteinander verglichen; alle gehören zu $x_0 = 0$. Je nach dem Vorzeichen der Geschwindigkeit ist die Steigung positiv (v_1) oder negativ (v_3). In allen drei Fällen ist die Geschwindigkeit konstant, die Beschleunigung ist Null.
In Abb. 2.1.2–5c werden Bahnkurven betrachtet, bei denen das Teilchen bei $x(0) = 0$ mit verschiedenen Geschwindigkeiten startet und konstant beschleunigt wird;

$x(t) = -gt^2/2 + vt$, $\dot{x}(t) = gt + v$, $\ddot{x}(t) = -g$.

Im Geschwindigkeits-Zeit-Diagramm erhält man parallele Kurven, die Änderung der Geschwindigkeit, die Beschleunigung, ist für alle Bahnen die gleiche, im betrachteten Falle $-g$.
Durchschnittsgeschwindigkeit. Sie ist eine pauschale Aussage, die aber in manchen Fällen vorrangig interessiert, z.B. wenn man mit dem Zug oder dem Auto an einen anderen Ort fährt. Die Formel für die Definition der Geschwindigkeit wird dabei mit gesamter Fahrstrecke und Gesamtfahrzeit ausgewertet.
Kreisbewegung (Abb. 2.1.2–3b). Bei ihr bestimmt $\varphi(t)$ die Art, wie der Kreis im Zeitablauf durchlaufen wird. Wächst der Winkel linear mit der Zeit an, $\varphi(t) = \omega t$, so spricht man von einer Bewegung mit konstanter *Winkelgeschwindigkeit* ω. Da der Winkel während eines Umlaufes, d.h. während der Umlaufzeit T, um 2π zunimmt, ist die Winkelgeschwindigkeit

$\omega = 2\pi/T = 2\pi f$.

Sie kann auch durch die *Frequenz* f, die Zahl der Umläufe pro Zeit (oft auch mit ν bezeichnet), ausgedrückt werden. Winkelgeschwindigkeit und Frequenz haben die Dimension $(\text{Zeit})^{-1}$. Bei der Kreisbewegung mit konstanter Winkelgeschwindigkeit ist der Betrag der Geschwindigkeit konstant. Da während der Umlaufszeit T die Strecke vom Kreisumfang $2\pi R$ zurückgelegt wird, ist sie $v = 2\pi R/T = \omega R$. Die Richtung der Geschwindigkeit ändert sich jedoch laufend. Detaillierte Aussagen erhält man aus der Bahnkurve $\boldsymbol{s}(t) = \boldsymbol{e}_x R \cos \omega t + \boldsymbol{e}_y R \sin \omega t$,
$\dot{\boldsymbol{s}}(t) = -\boldsymbol{e}_x \omega R \sin \omega t + \boldsymbol{e}_y \omega R \cos \omega t$,
$\ddot{\boldsymbol{s}}(t) = -\boldsymbol{e}_x \omega^2 R \cos \omega t - \boldsymbol{e}_y \omega^2 R \sin \omega t = -\omega^2 \boldsymbol{s}(t)$.
Der Zeitablauf der x- bzw. y-Komponente dieser Größen ist in Abb. 2.1.2–6 dargestellt. Aus dem Zusammenhang mit der Kreisbewegung wird verständlich, daß man eine Sinus- oder Cosinusabhängigkeit von der Zeit eine *rein periodische Bewegung* nennt.
Wegen der Richtungsänderung der Geschwindigkeit liegt eine beschleunigte Bewegung vor. Die Kreisbewegung erfordert die zum Zentrum des Kreises gerichtete Zentripedalbeschleunigung vom Betrag $\omega^2 R$. Man führt den *Vektor der Winkelgeschwindigkeit* $\boldsymbol{\omega} = \omega \boldsymbol{e}_z$ ein, der den Umlaufsinn als Rechtsschraube charakterisiert. Die Geschwindigkeit läßt sich als Vektorprodukt $\dot{\boldsymbol{s}} = \boldsymbol{\omega} \times \boldsymbol{s}$ schreiben.

2.1.3. Dynamik

Der Bewegungsablauf eines mechanischen Systems steckt in der Bahnkurve. Entsprechend der Anzahl der Freiheitsgrade des Systems umfaßt diese f Funktionen der Zeit. Um die Bahnkurve, ausgehend von bestimmten Anfangsbedingungen, berechnen zu können, muß man die Bewegungsgleichung kennen, die ein Erfahrungsgesetz ist.
Als Zusammenfassung des Erfahrungsmateriales und darüber hinausgehender Abstraktionen formulierte Newton 1687 in dem Buch „Philosophiae Naturalis Prinzipia Mathematica" seine *Axiome*:
(1) *Jeder Körper verbleibt im Zustand der Ruhe oder der gleichförmig geradlinigen Bewegung, wenn er nicht durch einwirkende Kräfte gezwungen wird, seinen Zustand zu ändern.*
Hier wird zum Ausdruck gebracht, daß der Impuls ($p = mv$) bei Abwesenheit von Kräften konstant ist. Das gleichberechtigte Auftreten von Ruhe und Bewegung mit konstanter

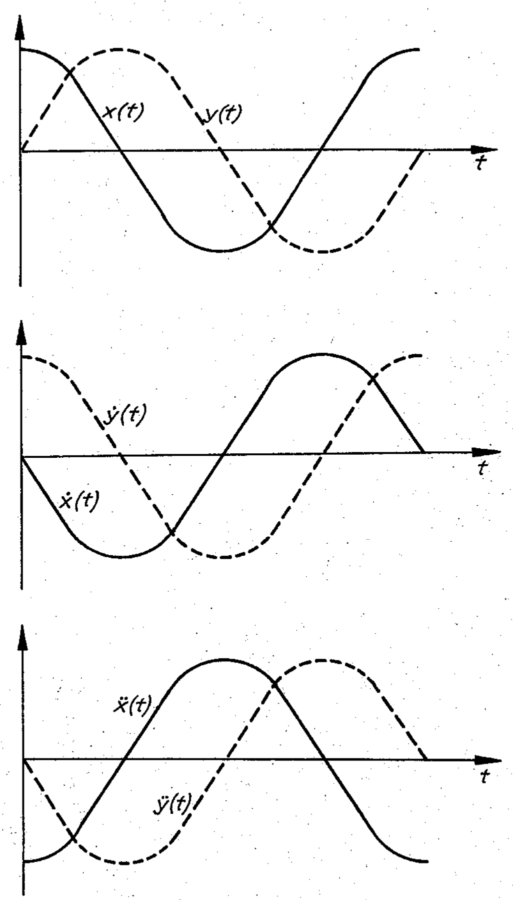


Abb. 2.1.2–6 Kreisbewegung mit konstanter Winkelgeschwindigkeit

Geschwindigkeit enthält das Galileische Relativitätsprinzip.

(2) *Die Änderung der Bewegung ist der Einwirkung der bewegenden Kraft proportional und erfolgt in Richtung der Kraft.*

Das ist die Aussage, die als Newtonsches Grundgesetz $\dot{p} = F$ bezeichnet wird.

(3) *Die Wirkung ist stets der Gegenwirkung gleich.*

Die Wirkung zweier Körper aufeinander ist betragsmäßig gleich groß und von entgegengesetzter Richtung (actio = reactio).

(4) *Kräfte addieren sich wie Vektoren.*

2.1.3.1. Newtonsches Grundgesetz. Die Masse kann in der klassischen Mechanik – im Unterschied zur relativistischen Mechanik – als konstant angesehen werden. Auch bei Raketen (vgl. 2.1.4.6.) ändert sich die Gesamtmasse, die die ausgestoßenen Gase mit umfaßt, zeitlich nicht. Das Newtonsche Grundgesetz für eine Punktmasse vereinfacht sich damit zu

$$m\ddot{s} = F, \quad m\ddot{x} = F.$$

Es ist eine Differentialgleichung für die Bahnkurve $s(t)$ oder $x(t)$. Es mußten viele Einzelerfahrungen gesammelt werden, und es erforderte ein hohes Abstraktionsvermögen, um diese Gleichung formulieren zu können. Insbesondere ist eine kräftefreie Bewegung nur im Gedankenexperiment realisierbar, weil tatsächlich immer Reibungskräfte auftreten.

Zur quantitativen Überprüfung der Grundgleichung der Mechanik kann man die *Atwoodsche Fallmaschine* benutzen. Zwei gleiche Massen sind durch einen Faden verbunden, der über einer Rolle hängt. Beide Massen verharren in Ruhe. Wird auf einer Seite eine Masse M' hinzugefügt, dann sinkt diese Seite nach unten, der Faden rollt ab. Die zurückgelegte Strecke x kann als Funktion der Zeit gemessen werden. Von der Masse des Fadens und der Rolle und von der Reibung soll abgesehen werden können. Die auf die beiden Massen M wirkende Schwerkraft trägt nicht zur Beschleunigung bei. Die Rolle trägt beide Massen. Die resultierende Kraft ist die auf M' wirkende Schwerkraft $F = M'g$. Insgesamt ist die Masse $m = 2M + M'$ in Bewegung zu versetzen. Man beobachtet $x \sim t^2$, also $\dot{x} = v \sim t$, also $\ddot{x} = a =$ konst. Es tritt eine konstante Beschleunigung a auf. Durch Variation von M und M' findet man $a \sim M'$ und $a \sim 1/(2M + M')$. Das heißt, die Beschleunigung ist der wirkenden Kraft F proportional und umgekehrt proportional zu der zu beschleunigenden Masse m. Das ist aber das oben genannte Grundgesetz. In 2.1.3.3. wird erläutert, daß die Kraft so definiert wird, daß der denkbare Proportionalitätsfaktor Eins ist.

Alle Erfahrungen bestätigen das an diesem Beispiel erläuterte Grundgesetz. Das gleiche Gesetz bestimmt die Pendelbewegung, die Bahn eines geladenen Teilchens in einem elektromagnetischen Feld oder die Planetenbewegung.

Da die Newtonsche Grundgleichung die Beschleunigung enthält, muß man Anfangsort und Anfangsgeschwindigkeit vorgeben, um die Bahnkurve berechnen zu können. Mit gegebenen *Anfangsbedingungen* folgt aus der Grundgleichung eine eindeutige Bahnkurve.

Bei *mehreren Massen* m_i erhält man ein gekoppeltes System von Differentialgleichungen für die Bahnen $s_i(t)$

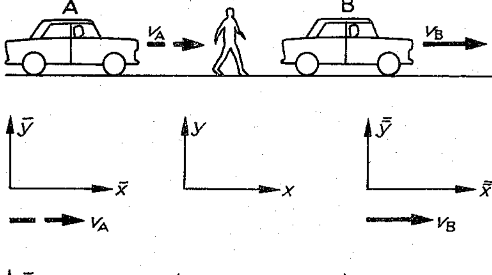
$$m_i\ddot{s}_i = f_i = f_a(s_i) + \sum_{j(\neq i)} f_{ij}(|s_i - s_j|).$$

Die Kraft f_i auf die Masse m_i unterteilt man in die *äußere Kraft* f_a, die nur vom Ort der einen Masse i abhängt, und die *paarweise Wechselwirkung* f_{ij}, durch die die Kopplung bewirkt wird. Wegen $f_{ij} = -f_{ji}$ (actio = reactio) fällt dieser Anteil bei Aufsummation heraus $M\ddot{s} = \sum_i f_a = F_a$, $M = \sum_i m_i$.

Ist $f_a =$ konst. oder für die Massenelemente eines ausgedehnten Körpers in guter Näherung durch den Wert des Kraftfeldes am Schwerpunkt ersetzbar, $f_a(s_i) \approx f_a(s)$, dann ist die Bewegungsgleichung des Schwerpunktes s abgekoppelt, und sie kann für sich gelöst werden. Dieser Sachverhalt rechtfertigt den Begriff der *Punktmasse*. Die Gesamtmasse M wird als im Schwerpunkt vereinigt gedacht. Die auf sie wirkende Kraft F_a ist die Summe aller an den Massenelementen angreifenden äußeren Kräfte.

2.1.3.2. Relativitätsprinzip. Die Erfahrungen zeigen, daß die *Absolutgeschwindigkeit* einer Masse nicht meßbar ist. Es lassen sich nur Geschwindigkeiten relativ zu einem anderen Körper messen.

Ein *Bezugssystem* wird benötigt, um die Lage, z.B. den Ortsvektor einer Punktmasse, zahlenmäßig angeben zu können. Es gehört dazu die Festlegung eines Koordinatenursprungs und der Richtung der Koordinatenachsen. In den meisten Fällen ist das Bezugssystem fest mit dem Labor verbunden. Es kann aber auch mit einem Raumschiff oder einem fahrenden Wagen verknüpft sein. Die oben genannte Erfahrung, daß man keine Absolutgeschwindigkeiten messen kann, ist in der Newtonschen Grundgleichung enthalten. Abb. 2.1.3–1 enthält ein Beispiel, wo die Fahrt zweier Au-



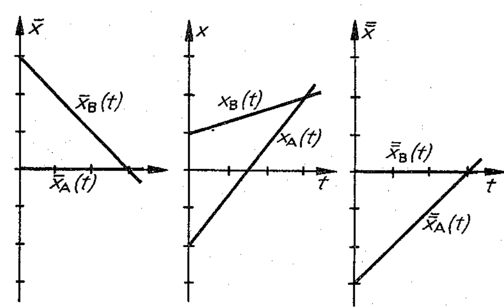


Abb. 2.1.3–1 Darstellung der Bewegung zweier Autos in drei verschiedenen Koordinatensystemen ($v_A = 4v_B$)

tos (A bzw. B) von drei verschiedenen Beobachtern verfolgt wird, den beiden Fahrern und einem Passanten. Die Bewegung wird in drei verschiedenen Bezugssystemen aufgezeichnet, die mit der Straße (x) bzw. den Autos ($\bar{x}$, $\bar{\bar{x}}$) verknüpft sind. Fahren die Autos auf der Straße mit den konstanten Geschwindigkeiten $v_A = 80$ km/h und $v_B = 20$ km/h, dann registrieren die Fahrer die Differenzgeschwindigkeiten ± 60 km/h. Die beobachteten Bahnkurven lauten

$x_A(t) = x_A(0) + v_A t \qquad x_B(t) = x_B(0) + v_B t$

$\bar{x}_A(t) = 0 \qquad \bar{x}_B(t) = x_B(0) - x_A(0) + (v_B - v_A)t$

$\bar{\bar{x}}_A(t) = x_A(0) - x_B(0) + (v_A - v_B)t \qquad \bar{\bar{x}}_B(t) = 0$

Es gilt $\ddot{x}_B = \ddot{\bar{\bar{x}}}_B = 0$. Eine kräftefreie Bewegung bleibt auch in einem relativ mit konstanter Geschwindigkeit bewegten Bezugssystem eine kräftefreie Bewegung. Die Newtonsche Bewegungsgleichung hat die gleiche Form in den drei Bezugssystemen, kein Bezugssystem ist ausgezeichnet. Auch wenn sich das Auto B beschleunigt bewegt, $x_B(t)$ eine beliebige Funktion der Zeit ist, dann ist die vom Fahrer A beobachtete Beschleunigung wegen $\bar{x}_B(t) = x_B(t) - v_A t - x_A(0)$ die gleiche wie die vom Passanten beobachtete $\ddot{\bar{x}}_B = \ddot{x}_B$.
Die Gesamtheit der relativ zueinander mit konstanter Geschwindigkeit bewegten Bezugssysteme nennt man *Inertialsysteme*.
In ihnen gilt die Newtonsche Grundgleichung $\dot{\boldsymbol{p}} = \boldsymbol{F}$.
Wie in dem in Abb. 2.1.3-1 dargestellten Beispiel ergibt sich die Geschwindigkeit eines Körpers beim Übergang zu einem anderen Inertialsystem als Summe, genauer Vektorsumme, der Einzelgeschwindigkeiten. Entsprechend der Relativgeschwindigkeit des Bezugssystems kann diese Summe beliebige Werte annehmen.
Galileisches Relativitätsprinzip. Damit bezeichnet man die Verknüpfung der Erfahrung, daß
(1) eine Absolutgeschwindigkeit nicht angebbar ist bzw. daß sich Inertialsysteme physikalisch nicht voneinander unterscheiden, mit der Aussage, daß
(2) die Geschwindigkeit beliebige, d. h. auch beliebig große, Werte annehmen kann.
Die Newtonsche Grundgleichung erfüllt dieses Relativitätsprinzip. Dieses Prinzip hätte eine geringe Aussagekraft, wenn es mit nichtmechanischen Mitteln gelänge, eine Absolutgeschwindigkeit zu messen. Das Umgekehrte ist der Fall. Alle physikalischen Erfahrungen zeigen, daß eine Absolutgeschwindigkeit nicht bestimmbar ist. Alle physikalischen Erscheinungen genügen dem gleichen Relativitätsprinzip.
Ein solch allgemeingültiges Prinzip ist eine Erfahrung, auf die man beim Vorstoß in neue Gebiete aufbaut. Man wird

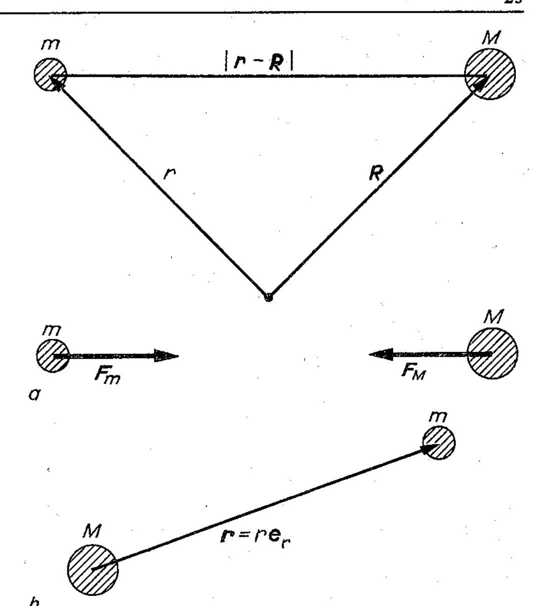
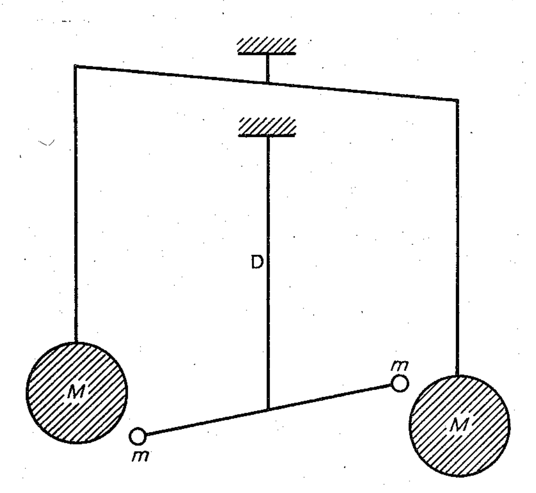


Abb. 2.1.3-2 (a) vektorielle Darstellung der Gravitationskraft und (b) Beschreibung mit dem Koordinatenursprung in M

zunächst nur solche Theorien ins Auge fassen, die diesem Prinzip genügen. Aufgrund der Erfahrungen muß das Galileische Relativitätsprinzip in seiner Aussage über die Addition der Geschwindigkeiten für hohe Geschwindigkeiten abgeändert werden. Die Maximalgeschwindigkeit ist die Vakuumlichtgeschwindigkeit (vgl. 2.5.2.1.).

2.1.3.3. Kräfte in der Natur. In der Newtonschen Grundgleichung taucht die physikalische Größe Kraft auf. Über diese Gleichung ist also die Dimension der Kraft bestimmt. Die Kraft ist eine abgeleitete Größe. Ihre Einheit ist das Newton ($1\,\mathrm{N} = 1\,\mathrm{kgms^{-2}}$). In der Regel wird eine Kraft nicht



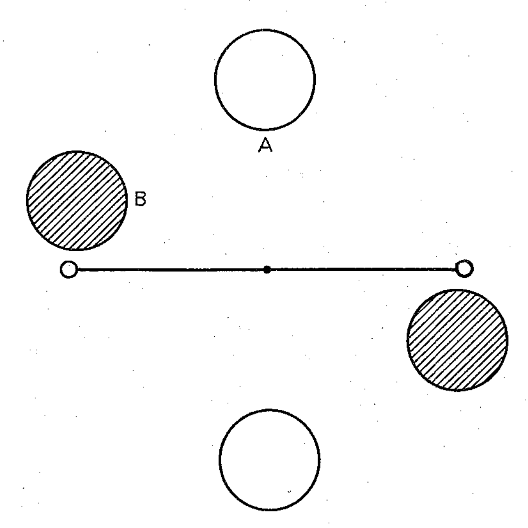


Abb. 2.1.3-3 Drehwaage von Cavendish, rechts Draufsicht

durch Verwendung der Newtonschen Grundgleichung gemessen, sondern durch Vergleich mit schon bestimmten Kräften, der Gravitationskraft oder der Federkraft.

Gravitationskraft. Als solche bezeichnet man die Anziehung zwischen zwei Massen m, M. Sie ist umgekehrt proportional zu deren Abstandsquadrat. Die vektorielle Form der Kraft auf die Masse m ist

$$\boldsymbol{F}_m = -G\frac{mM}{|\boldsymbol{r}-\boldsymbol{R}|^2}\,\frac{\boldsymbol{r}-\boldsymbol{R}}{|\boldsymbol{r}-\boldsymbol{R}|}$$

(Abb. 2.1.3–2a). Die Kraft auf die Masse M ist entgegengesetzt gleich groß. Man erhält sie, wenn man $\boldsymbol{r}$ bzw. m mit $\boldsymbol{R}$ bzw. M vertauscht,

$$\boldsymbol{F}_M = -G\frac{Mm}{|\boldsymbol{R}-\boldsymbol{r}|^2}\,\frac{\boldsymbol{R}-\boldsymbol{r}}{|\boldsymbol{R}-\boldsymbol{r}|} = -\boldsymbol{F}_m.$$

Legt man den Koordinatenursprung in die Masse M, dann vereinfacht sich die Formel für die Kraft auf die Masse m zu

$$\boldsymbol{F} = -\boldsymbol{e}_r GmM/r^2.$$

(Abb. 2.1.3–2b). Ist M eine ausgedehnte kugelsymmetrische Massenverteilung, dann ist die Kraft, die auf eine sich außerhalb der Verteilung befindliche Masse m wirkt, die gleiche wie bei einer Punktmasse M.

Der Proportionalitätsfaktor $G = 6{,}67 \cdot 10^{-11}\,\mathrm{kg^{-1}m^3s^{-2}}$, die *Gravitationskonstante*, muß experimentell ermittelt werden. Cavendish verwendete eine Drehwaage (Abb. 2.1.3–3). Zwei Massen $m = 730$ g waren an den Enden eines 2 m langen Holzstabes befestigt und an einem Draht D aufgehängt. Nach der Justierung, bei der sich die großen Massen $M = 158$ kg in der Position A befanden, wurden diese Massen in die Position B gebracht. Die Massen m wurden durch die Gravitation ausgelenkt, bis die Gravitationskraft durch die Verdrillung des Drahtes kompensiert wurde. Es gab auch andere Messungen, bei denen der Zeitablauf der Auslenkung verfolgt wurde.

Schwerkraft ist die Gravitationskraft, die die Erde (Masse M_E, Radius R) auf eine Masse in der Nähe (Höhe z) der Erdoberfläche ausübt. Für $z \ll R$ erhält man die von der Höhe z unabhängige Kraft

$$\boldsymbol{F} = -G\frac{mM_\mathrm{E}}{(R+z)^2}\,\boldsymbol{e}_z \approx -G\frac{M_\mathrm{E}}{R^2}\,m\boldsymbol{e}_z \equiv -mg\boldsymbol{e}_z = m\boldsymbol{g}.$$

Die *Schwerebeschleunigung* $g = GM_\mathrm{E}/R^2$ hat den mittleren Wert $g = 9{,}81\,\mathrm{ms^{-2}}$. Durch die Drehung der Erde und durch die dadurch bewirkte Abplattung der Erde (vgl. 2.1.9.) ist die beobachtete Kraft vom Breitengrad abhängig $g_\varphi = (9{,}832 - 0{,}052\cos^2\varphi)\,\mathrm{ms^{-2}}$. Für $\varphi = 45°$ erhält man daraus $g_{45°} = 9{,}806\,\mathrm{ms^{-2}}$. Aus dem Wert von g kann man bei bekannten Werten R und G auf M_E schließen.

Die Erde übt auf eine Masse von 1 kg die Schwerkraft 9,81 N aus. Im Vergleich dazu ist die Kraft der Sonne auf die gleiche Masse im Abstand des Erdbahnradius etwa 3 mN, die des Mondes im Abstand Mond-Erde 0,03 mN.

Coulomb-Kraft ist die Kraft zwischen zwei Ladungen, z. B. die Kraft zwischen Proton und Elektron. Diese Kraft bestimmt den Atombau. Aber auch makroskopische Körper können durch Mangel oder Überschuß von Elektronen geladen sein, wie etwa die Platten eines Kondensators. Die Coulomb-Kraft hat eine der Gravitationskraft ähnliche Form, statt der Massen m, M treten die Ladungen q, Q auf,

$$\boldsymbol{F}_q = \frac{qQ}{4\pi\varepsilon_0|\boldsymbol{r}-\boldsymbol{R}|^2}\,\frac{\boldsymbol{r}-\boldsymbol{R}}{|\boldsymbol{r}-\boldsymbol{R}|},\quad \boldsymbol{F}_Q = -\boldsymbol{F}_q.$$

Ein entscheidender Unterschied zur Gravitationskraft besteht darin, daß die Coulomb-Kraft sowohl anziehend wie auch abstoßend sein kann, da es positive und negative Ladungen gibt. Die Dielektrizitätskonstante des Vakuums hat den Wert $\varepsilon_0 = 8{,}8542 \cdot 10^{-12}\,\mathrm{As/Vm}$. Der Faktor $1/4\pi\varepsilon_0 \approx 9 \cdot 10^9\,\mathrm{Vm/As}$ hat einen leichter einprägsamen Wert.

Die Coulomb-Kraft zwischen zwei Elektronen ist wesentlich größer als deren Massenanziehung, wie das Verhältnis $F_\mathrm{Grav.}/F_\mathrm{Coul.} \approx 10^{-42}$ ausweist. Bei großen Massenansammlungen wird dessenungeachtet die Gravitation dominierend, da sich die Gravitationskräfte immer addieren. Dagegen neutralisieren sich in allen Körpern die positiven und negativen Ladungen im wesentlichen.

Die Coulomb-Kraft enthält mit q eine Eigenschaft der Ladung, an der die Kraftwirkung beobachtet wird. Über $\boldsymbol{F} = q\boldsymbol{E}$ definiert man das elektrische Feld der Ladung Q. Erzeugt eine beliebige Ladungsverteilung das elektrische Feld $\boldsymbol{E}$, dann beobachtet man die Kraft $\boldsymbol{F} = q\boldsymbol{E}$. Dies ist also der Ausdruck für die *Kraft auf eine Ladung im elektrischen Feld.*

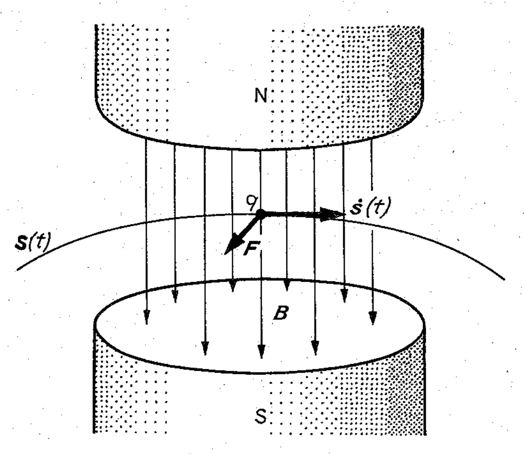


Abb. 2.1.3–4 Kraft im Magnetfeld ($q < 0$)

Auf eine bewegte Ladung wirkt auch eine *Kraft im Magnetfeld* $\boldsymbol{F} = q\dot{\boldsymbol{s}} \times \boldsymbol{B}$. $\boldsymbol{B}$ ist das Feld der magnetischen Induktion. Die Richtungsverhältnisse werden in Abb. 2.1.3–4 veranschaulicht. Die *Lorentz-Kraft*, die Kraft auf eine Ladung im elektromagnetischen Feld, ist also

$$\boldsymbol{F} = q(\boldsymbol{E} + \dot{\boldsymbol{s}} \times \boldsymbol{B}).$$

Elastische Kräfte treten z. B. bei der Dehnung einer Feder, bei der Biegung eines Balkens oder bei der Verdrillung eines Drahtes auf. Sie wirken der Verformung entgegen. Bei kleinen Verformungen sind es rücktreibende Kräfte proportional zur Deformation. Diese Kräfte haben ihren Ursprung in der elektrischen Wechselwirkung zwischen den atomaren Bausteinen. Zu ihrer Charakterisierung genügt jedoch die Angabe des Proportionalitätsfaktors. Dieser wird als Materialkonstante, z. B. elastische Konstanten, experimentell ermittelt.

Federkraft. Die Federkraft einer Feder mit vielen Windungen

$$F = -kx$$

ist auch bei relativ großen Auslenkungen noch proportional zur Dehnung. Eine Feder ist daher gut zur Messung von Kräften geeignet.

Reibung. Hierbei summieren sich viele atomare Wechsel-

Tab. 2.1.3–5 Koeffizienten der Haftreibung (μ_H), der Gleitreibung (μ_G) und der rollenden Reibung (f bzw. $\mu_R = f/r$)

Stahl auf Stahl	$\mu_H = 0{,}15$	$\mu_G = 0{,}05$
Stahl auf Eis	$\mu_H = 0{,}027$	$\mu_G = 0{,}014$
Holz auf Stein	$\mu_H = 0{,}7$	$\mu_G = 0{,}3$
Stahlräder auf Schienen	$f = 0{,}0001$ m	
Kugellager	$f = 0{,}001$ cm	
Fuhrwerk mit eisernen Radreifen		
auf Asphalt	$\mu_R = 0{,}01$	
auf Steinpflaster	$\mu_R = 0{,}02$	
im Sand	$\mu_R = 0{,}2$	

wirkungen zu einem pauschalen Effekt. Der genaue Wert der Reibungskraft hängt stark von der Oberflächenbeschaffenheit ab, die insbesondere durch Schmierstoffe beeinflußt werden kann. Die bei der *Gleitreibung* auftretende Reibungskraft ist der Kraft auf die Unterlage proportional. Sie wirkt tangential zur Oberfläche der Bewegungsrichtung entgegengerichtet. Bei einer waagerechten Oberfläche ist der Betrag der Reibungskraft $F_R = \mu Mg$. Bevor man einen ruhenden Körper in Bewegung versetzen kann, muß man die *Haftreibung* überwinden, die größer als die Gleitreibung ist. Die *Reibung beim Rollen* ist geringer als die Gleitreibung. Sie hängt vom Radius des Rades ab und wird oft in der Form $F_R = (f/r)Mg$ angegeben (Tab. 2.1.3–5). Bei einer Lokomotive ist für die Zugkraft die Haftreibung entscheidend. Bei den Eisenbahnwagen ist zwischen Rädern und Schienen Rollreibung und in den Achsen Gleitreibung zu überwinden.

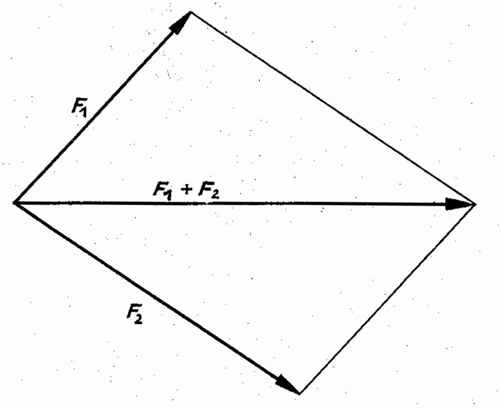


Abb. 2.1.3–6 Vektorielle Addition von Kräften

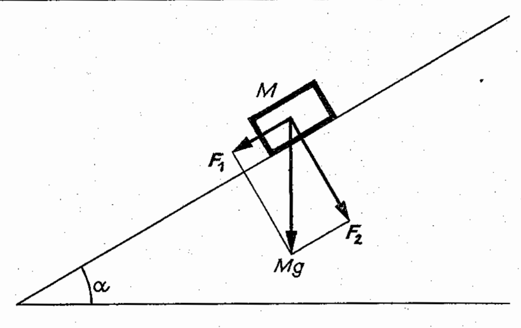
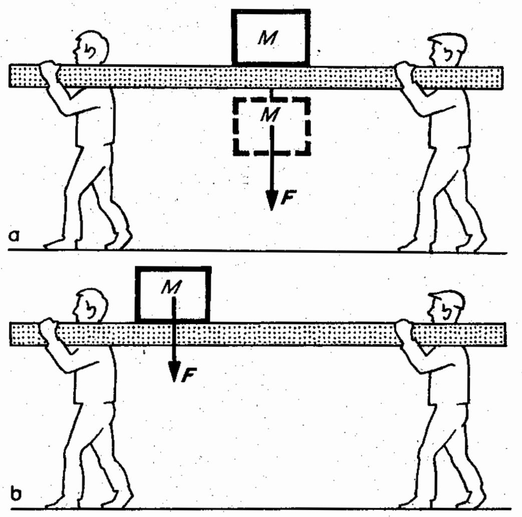


Abb. 2.1.3–7 Zerlegung der Schwerkraft für eine Masse auf einer schiefen Ebene; M = Masse, F_1 = Hangabtriebskraft, F_2 = Normalkraft

Die *Luftreibung* oder die Reibung bei der Bewegung in einer Flüssigkeit ist in einem weiten Geschwindigkeitsbereich proportional zur Geschwindigkeit. So beschreibt die *Stokessche Reibungskraft* $F_R = 6\pi\eta Rv$ (Zähigkeit η) die Kraft auf eine Kugel vom Radius R. Die Reibungskraft, die auf eine Masse m wirkt, schreibt man oft zweckmäßig in der Form $F_R = -\gamma m\dot{s}$, worin γ ein Reibungskoeffizient der Dimension (Zeit)$^{-1}$ ist.



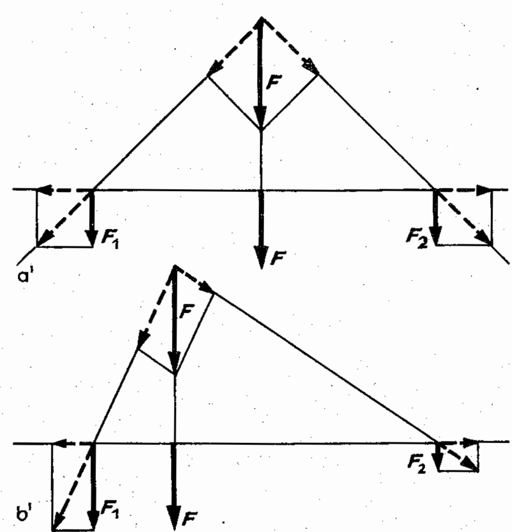


Abb. 2.1.3–8 (a) Für beide Positionen der Last M ist die Belastung der Träger gleich groß. (b) Ein Verschieben von F quer zur Wirkungslinie verändert die Belastung. Rechts ist die Belastung der einzelnen Träger konstruiert

2.1.3.4. Zusammensetzen von Kräften. Eine Kraft ist ein Vektor, der durch Betrag und Richtung charakterisiert ist. Greifen an einer Masse mehrere Kräfte an, so ist die resultierende Kraft durch die Vektorsumme, das *Kräfteparallelogramm* in Abb. 2.1.3–6, gegeben (4. Newtonsches Axiom). Umgekehrt kann man eine Kraft in Komponenten zerlegen. Bewegt sich ein Körper auf einer *schiefen Ebene* (Abb. 2.1.3–7), so zerlegt man die Schwerkraft zweckmäßigerweise in eine Komponente $F_1 = Mg\sin\alpha$ parallel zur Ebene (Hangabtriebskraft) und eine Komponente $F_2 = Mg\cos\alpha$ senkrecht zur Ebene. F_2 bestimmt die auftretende Reibungskraft (μF_2), F_1 die Beschleunigung der Masse.
Eine Kraft kann man längs ihrer Wirkungslinie verschieben, ohne ihre Wirkung zu verändern. Abb. 2.1.3–8 veranschaulicht, wie man parallele Kräfte addiert. Abb. 2.1.3–9 zeigt die Kräfte, die die Türangeln übertragen müssen, um das Gewicht der Tür zu kompensieren. Die obere Angel wird auf Zug $\boldsymbol{F}_Z$, die untere durch Druck $\boldsymbol{F}_D$ beansprucht. Wie sich die vertikalen Kräfte auf die beiden Angeln verteilen, hängt von der genauen Auflage ab und kann z. B. durch eine Unterlegscheibe verändert werden.
Durch eine Rolle wird eine Kraft umgelenkt. Abb. 2.1.3–10a zeigt die Beanspruchung der Halterung einer *festen Rolle*. Bei einer *losen Rolle* (Abb. 2.1.3–10b) genügt eine Kraft vom Betrag des halben Gewichtes, um eine Last zu halten. Der *Flaschenzug* ist eine Kombination loser und fester Rollen (Abb. 2.1.3–10c).
Ein wichtiger, im Zusammenhang mit Kräften auftretender Begriff ist das *Drehmoment* $\boldsymbol{M}$. Das Drehmoment einer Kraft, die in einem Punkt $\boldsymbol{r}$ angreift, bezogen auf den Koordinatensprung, ist $\boldsymbol{M} = \boldsymbol{r} \times \boldsymbol{F}$. Der Betrag des Drehmomentes $M = aF$ ist gleich dem Produkt aus dem Betrag der Kraft F und dem Abstand a der Wirkungslinie vom Bezugspunkt. Der Vektor Drehmoment $\boldsymbol{M}$ steht senkrecht auf der von den Vektoren $\boldsymbol{r}$ und $\boldsymbol{F}$ aufgespannten Ebene. Diese Eigenschaften stecken zusammengefaßt in dem Vektorprodukt $\boldsymbol{r} \times \boldsymbol{F}$. Der Wert des Drehmomentes ist also von der Wahl des Koordinatensprunges abhängig. Das Drehmoment eines *Kräftepaares* (Abb. 2.1.3–11a) $\boldsymbol{M} = \boldsymbol{r}_1 \times \boldsymbol{F} + \boldsymbol{r}_2 \times (-\boldsymbol{F}) = (\boldsymbol{r}_1 - \boldsymbol{r}_2) \times \boldsymbol{F}$ ist jedoch von der Wahl des Bezugspunktes unabhängig. Sein Betrag ist gleich dem Produkt von Kraft mal dem senkrechten Abstand des Kräftepaares. Da in Abb. 2.1.3–9 die Wirkungslinie der Gewichtskraft der Tür nicht durch die Angeln verläuft, müssen diese noch ein Drehmoment kompensieren. In Abb. 2.1.3–11b bewirkt der Eimer ein Drehmoment R_1F_1 bezüglich der Achse. Um den Eimer zu halten, muß ein gleich großes Drehmoment R_2F_2 wirken, wozu bei $R_2 > R_1$ eine kleinere Kraft benötigt wird.
Das Abbremsen eines Körpers erfordert entsprechend der Newtonschen Bewegungsgleichung eine im Schwerpunkt angreifende, der Bewegungsrichtung entgegengesetzte Kraft. Wird das Abbremsen wie beim Fahrrad oder beim Auto durch die Reibung der Räder auf der Straße bewirkt, so erzeugt die Reibungskraft neben der im Schwerpunkt angreifenden, bremsenden Kraft noch ein Drehmoment. Bei

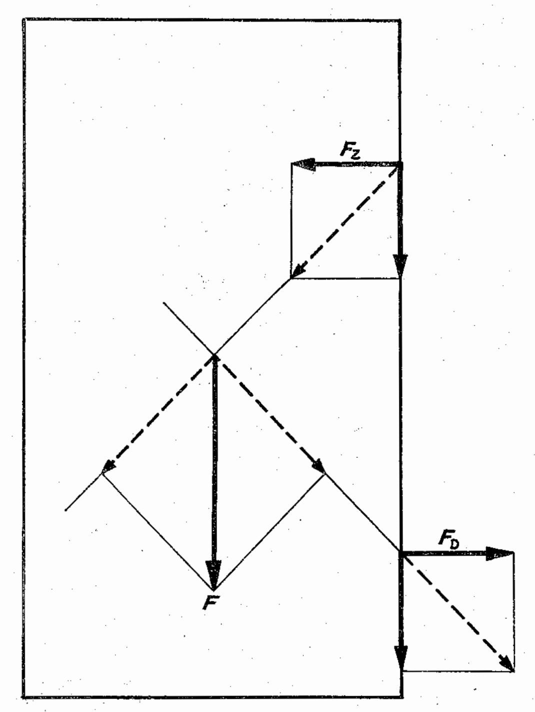


Abb. 2.1.3–9 Kräfte an einer Tür

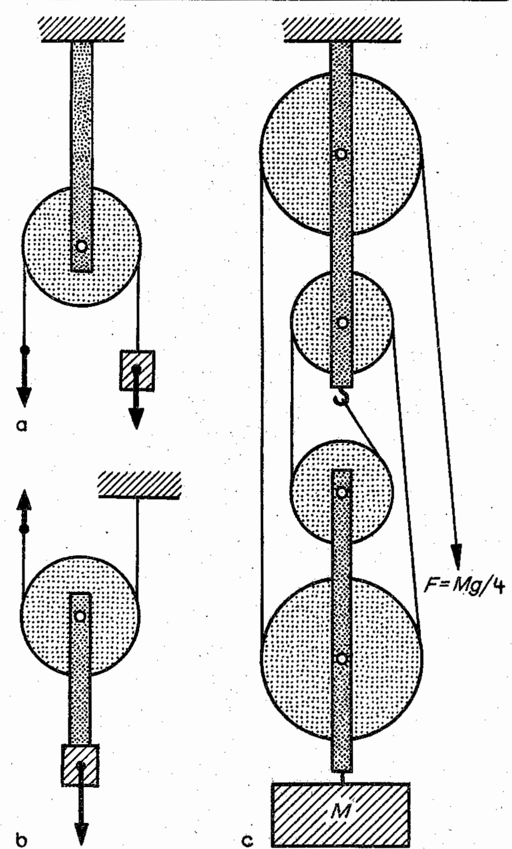


Abb. 2.1.3–10 (a) feste, (b) lose Rolle und (c) Flaschenzug

Vorderradbremsung erhöht dieses die Bremswirkung, jedoch kann sich dabei das Fahrzeug überschlagen.

2.1.3.5. Arbeit und Leistung. Arbeit. Beim Heben einer Last oder beim Dehnen eines Expanders wird Arbeit geleistet, nicht dagegen bei bloßem Halten eines Gewichtes. Als Arbeit wird das Produkt von zurückgelegter Wegstrecke Δs und der Komponente der äußeren Kraft in Richtung des Weges definiert, $W = \Delta s F_a$. Beim Heben muß die äußere Kraft die Schwerkraft kompensieren bzw. mindestens um einen infinitesimalen Betrag größer sein. Wird die Kraft über dem Weg aufgetragen (Abb. 2.1.3–12), dann läßt sich $\Delta s F_a$ durch ein Flächenelement veranschaulichen. Die *Hubarbeit* $W = mgh$ ist das Produkt von Schwerkraft und Hubhöhe h. In Abb. 2.1.3–12a ist es die gesamte schraffierte Fläche.

Beim Heben mittels einer *losen Rolle* (vgl. Abb. 2.1.3–10b) ist die äußere Kraft gleich dem halben Gewicht $F_a = mg/2$. Jedoch muß das Seil um die Strecke $\Delta s = 2h$ nach oben gezogen werden, wenn die Masse um h angehoben werden soll. Die benötigte Arbeit $\Delta s F_a = mgh$ ist also unabhängig von der Art des Anhebens. Mit einem *Flaschenzug* (vgl. Abb. 2.1.3–10c) kann die erforderliche äußere Kraft weiter verringert werden. In ähnlicher Weise dienen ein *Hebel* oder eine *Kurbel* (vgl. 2.1.3–11) dazu, eine Arbeit durch verringerte Kraft aber verlängerten Weg zu leisten und damit z. B. dazu, schwere Lasten anzuheben. Der Betrag der Arbeit kann dadurch nicht verändert werden, wie auch das *Hebelgesetz* „Kraft mal Kraftarm = Last mal Lastarm" zum Ausdruck bringt.

Ändert sich die Kraft mit dem Ort, wie beim Dehnen einer Feder (Abb. 2.1.3–12b), dann muß der Flächeninhalt i. allg. durch Integration berechnet werden. Bei der proportional zur Auslenkung anwachsenden Federkraft entspricht die Arbeit der äußeren Kraft bei Dehnung um L der Dreiecksfläche, $W = aL^2/2$. Die Kraftkomponente senkrecht zur Bewegungsrichtung leistet keine Arbeit. Verwendet man den vektoriellen Charakter von Verschiebung und Kraft, dann ist die Arbeit durch

$$dW = d\boldsymbol{s} \cdot \boldsymbol{F}_a, \quad W = \int_{s_0}^{s_1} d\boldsymbol{s} \cdot \boldsymbol{F}_a(s)$$

gegeben. Die Arbeit hat die *Einheit* Newtonmeter (Nm) oder Joule (J), $1\,J = 1\,Nm = 1\,kgm^2s^{-2}$. Wird an einem System Arbeit geleistet, so wird ihm Energie zugeführt. Diese Energiezufuhr bewirkt eine Erhöhung der kinetischen Energie (Beschleunigung, Erhöhung der Geschwindigkeit) oder der potentiellen Energie (Hubarbeit, Dehnen einer Feder), oder die Arbeit wird in nichtmechanische Energie, z. B. Wärmeenergie (Reibung) oder elektrische Energie (Dynamo) umgewandelt.

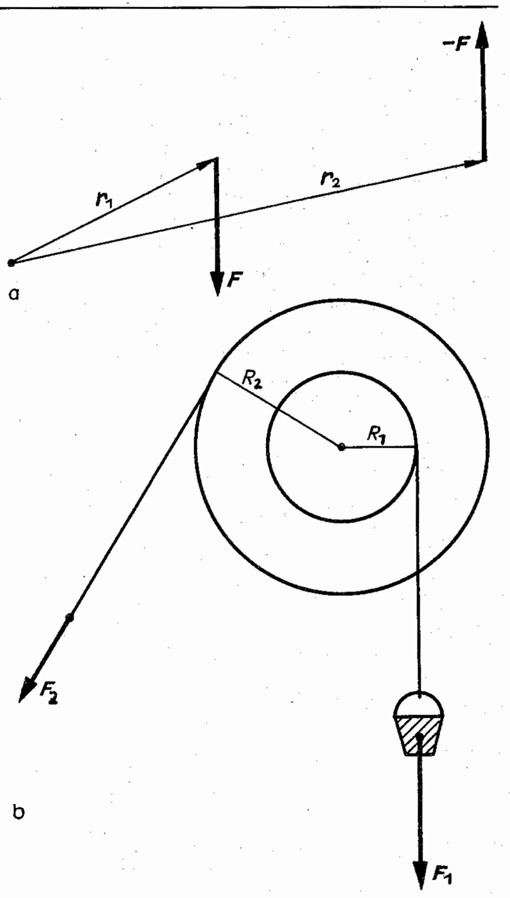


Abb. 2.1.3–11 Drehmoment (a) eines Kräftepaares und (b) bei einer Winde

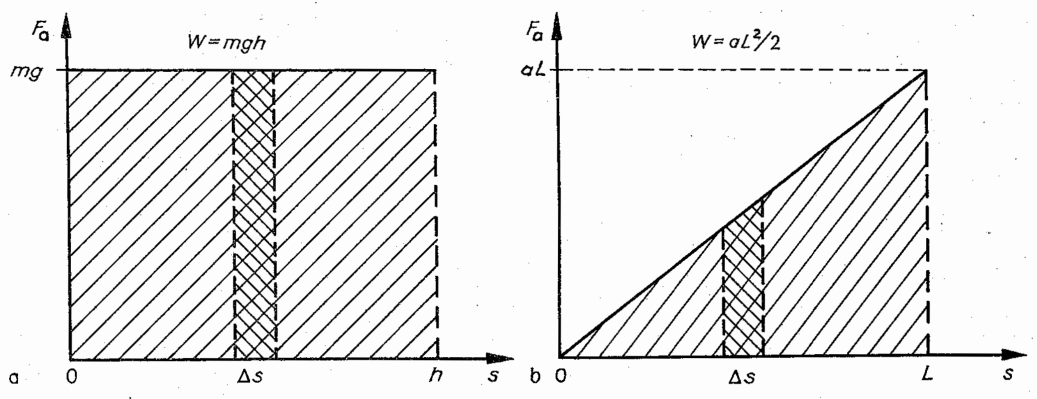


Abb. 2.1.3–12 (a) Hubarbeit und (b) Arbeit beim Dehnen einer Feder

Leistung beschreibt, in welcher Zeit eine Arbeit vollbracht wird. Wird ein Gewicht in der Zeit Δt um die Strecke Δs angehoben, dann ist die Leistung $P = F_a \Delta s/\Delta t = F_a v$. Sie ist durch das Produkt $P = F_a \cdot \dot{s}$ von äußerer Kraft und Geschwindigkeit gegeben, wobei in das Skalarprodukt wiederum nur die Komponente der Kraft in Richtung des Weges eingeht. Die Einheit der Leistung ist Joule/Sekunde oder Watt (W), $1\,W = 1\,J/s = 1\,Nm/s = 1\,kgm^2s^{-1}$. Man muß klar auf den Unterschied zwischen Arbeit und Leistung achten, weil in der Umgangssprache beide Begriffe vermischt werden, statt Arbeit wird vielfach Leistung gesagt.

2.1.3.6. Vektorfelder. Ein Feld ordnet in der Physik jedem Raumpunkt $\boldsymbol{r}$ einen Wert einer physikalischen Größe zu. Das Temperaturfeld $T(\boldsymbol{r})$ gibt die Temperaturverteilung an, das elektrische Feld $\boldsymbol{E}(\boldsymbol{r})$ die elektrische Feldstärke, das Kraftfeld $\boldsymbol{F}(\boldsymbol{r})$ den Vektor der Kraft, die auf die betrachtete Masse am Ort $\boldsymbol{r}$ wirkt.

Bei einem *skalaren Feld* ist die physikalische Größe ein Skalar: Temperatur, Dichte, Höhe über N.N., bei einem *Vektorfeld* ist die physikalische Größe ein Vektor: Kraftfeld, elektrisches Feld, magnetisches Feld, Geschwindigkeitsfeld.

Zur Veranschaulichung eines skalaren Feldes $V(\boldsymbol{r})$ verwendet man das *Höhenlinienbild*, wie es von Landkarten bekannt ist. Punkte, an denen die physikalische Größe gleiche Werte aufweist, sind miteinander verbunden. Bei einer ebenen Verteilung ergeben sich Linien, bei einer räumlichen Verteilung Flächen $V(\boldsymbol{r}) = \text{konst.}$

Zur Veranschaulichung eines Vektorfeldes verwendet man das Pfeilbild oder das Stromlinienbild. Abb. 2.1.3–13 zeigt qualitativ die Geschwindigkeitsverteilung des Wassers in einem Fluß, wobei in der Flußmitte die Geschwindigkeit

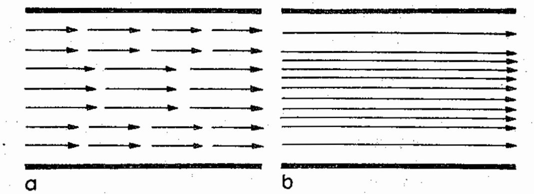


Abb. 2.1.3–13 (a) Pfeilbild und (b) Stromlinienbild

am größten ist. Bei dem *Pfeilbild* wird das Vektorfeld an einigen Punkten entsprechend Richtung und Größe eingetragen. Beim *Stromlinienbild* ist die Richtung des Vektorfeldes durch die Tangente an die Stromlinie gegeben. Der Betrag wird durch die Dichte der Stromlinien charakterisiert. Dort, wo die Stromlinien dichter beieinander liegen, ist der Betrag größer. Abb. 2.1.3–14 veranschaulicht die laminare Strömung durch ein Rohr mit wechselndem Querschnitt. Das Pfeilbild ist dem Ungeübten leichter lesbar, das Stromlinienbild ist aussagekräftiger.

Die charakteristischen Eigenschaften eines Vektorfeldes

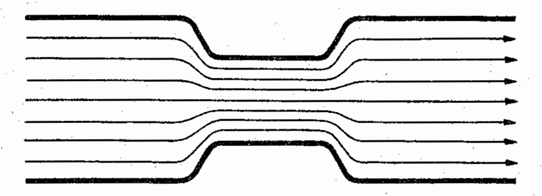


Abb. 2.1.3–14 Geschwindigkeitsfeld im Rohr

$\boldsymbol{F}(\boldsymbol{r})$ sind seine *Quellen* Q (oder Senken $Q < 0$) und *Wirbel* W. Von einer Quelle gehen Stromlinien aus. Die Quellstärke ist durch das Oberflächenintegral

$$Q = \oint\!\!\oint \mathrm{d}\boldsymbol{a} \cdot \boldsymbol{F}, \quad \varrho = \operatorname{div} \boldsymbol{F} = \nabla \cdot \boldsymbol{F},$$

gegeben, die Quelldichte ϱ durch die Divergenz (vgl. 13.3.6.2.). ∇ ist der Vektordifferentialoperator, $\nabla = \boldsymbol{e}_x \partial/\partial x + \boldsymbol{e}_y \partial/\partial y + \boldsymbol{e}_z \partial/\partial z$. Ein Vergleich mit dem Feld der Gravitationskraft zeigt, daß die Massen die Quellen der Gravitationskraft sind. Bei Wirbeln liegen im einfachsten Fall geschlossene Feldlinien um den Wirbel. Die Wirbelstärke ist durch das Wegintegral

$$W = \oint \mathrm{d}\boldsymbol{r} \cdot \boldsymbol{F}, \quad \boldsymbol{j} = \operatorname{rot} \boldsymbol{F} = \nabla \times \boldsymbol{F},$$

bestimmt, die Wirbeldichte $\boldsymbol{j}$ durch die Rotation (vgl. 13.3.6.2.). So sind z.B. die elektrischen Ströme Wirbel des Magnetfeldes, die elektrische Stromdichte ist die Wirbeldichte.

Ein Vektorfeld ist bestimmt, wenn man seine Quellen und Wirbel und sein Verhalten auf dem Rand des interessierenden Gebietes kennt.

Für die Betrachtung ist es vereinfachend, daß viele wichtige Kraftfelder wirbelfrei sind. Ein *wirbelfreies Feld*, d. h. ein Feld, dessen Rotation verschwindet, läßt sich aber als *Gradient eines skalaren Feldes* schreiben: $\boldsymbol{F}(\boldsymbol{r}) = -\operatorname{grad} V(\boldsymbol{r}) = -\nabla V(\boldsymbol{r})$ für $\nabla \times \boldsymbol{F} = 0$ bzw. $\oint \mathrm{d}\boldsymbol{r} \cdot \boldsymbol{F} = 0$. $V(\boldsymbol{r})$ nennt man das *Potential* des Feldes $\boldsymbol{F}(\boldsymbol{r})$. Damit eröffnet sich eine vereinfachte Veranschaulichung des Vektorfeldes über die *Potentialflächen* $V(\boldsymbol{r}) = \text{konst.}$ Das Vektorfeld zeigt überall in Richtung der Normalen der Potentialfläche. Für einige Kraftfelder werden die Potentiale zusammengestellt:

Gravitationskraft	$\boldsymbol{F}(\boldsymbol{r}) = -G\dfrac{mM}{r^2}\boldsymbol{e}_r,$	$V(\boldsymbol{r}) = -G\dfrac{mM}{r},$
Schwerkraft	$\boldsymbol{F} = m\boldsymbol{g},$	$V(\boldsymbol{r}) = -m\boldsymbol{g}(\boldsymbol{r} - \boldsymbol{r}_0),$
Coulomb-Kraft	$\boldsymbol{F} = \dfrac{qQ}{4\pi\varepsilon_0 r^2}\boldsymbol{e}_r,$	$V(\boldsymbol{r}) = \dfrac{qQ}{4\pi\varepsilon_0 r},$
Federkraft	$\boldsymbol{F} = k\boldsymbol{x},$	$V(x) = kx^2/2.$

In einem Kraftfeld mit dem Potential $V(x)$ zeigt die Kraft $\boldsymbol{F} = -\nabla V(x) = -\boldsymbol{e}_x \mathrm{d}V/\mathrm{d}x \equiv F(x)\boldsymbol{e}_x$ in x-Richtung. Das Vorzeichen und die Größe von $F(x)$ kann aus dem Verlauf der Kurve $V(x)$ abgelesen werden (Abb. 2.1.3–15a). Für $x < x_0$ und $x > x_1$ ist die Kraft abstoßend (von $x = 0$ aus gesehen), für $x_0 < x < x_1$ anziehend. Bei x_0 und x_1 verschwindet die Kraft, wobei x_0 eine stabile und x_1 eine labile Gleichgewichtslage ist.

2.1.3.7. Erhaltungsgrößen. In dem Newtonschen Grundgesetz $m\ddot{\boldsymbol{s}} = \boldsymbol{F}$ steht der Wert des Kraftfeldes $\boldsymbol{F}(\boldsymbol{r})$ am Ort der Punktmasse $\boldsymbol{F}[\boldsymbol{s}(t)]$. Im allgemeineren Fall kann das Kraftfeld auch von der Geschwindigkeit und explizit von der Zeit abhängen, $\boldsymbol{F}[\boldsymbol{s}(t), \dot{\boldsymbol{s}}(t), t]$. In der Mechanik wird untersucht, welche Bahnkurven in einem gegebenen Kraftfeld auftreten können. Diese Mannigfaltigkeit läßt sich am übersichtlichsten diskutieren, wenn man Erhaltungsgrößen aufsucht.

Eine *Erhaltungsgröße* ist eine physikalische Größe A, die eine Funktion von Ort und Geschwindigkeit der betrachteten Masse ist, deren Wert sich aber im Zeitablauf nicht ändert,

$$A = A[\boldsymbol{s}(t), \dot{\boldsymbol{s}}(t)] \text{ mit } \mathrm{d}A/\mathrm{d}t = 0.$$

Kennt man eine Erhaltungsgröße, so hat man also eine für alle Lösungen der Bewegungsgleichung geltende Aussage. Die Form des Kraftfeldes bestimmt im wesentlichen, welche Erhaltungsgrößen auftreten.

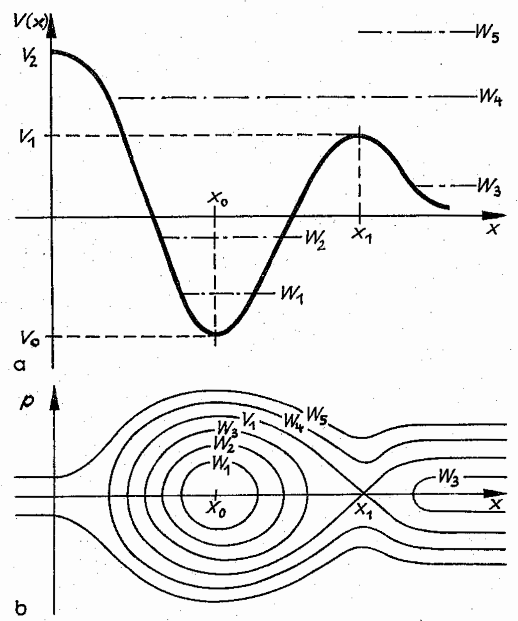


Abb. 2.1.3–15 (a) Potentialkurve und (b) zu verschiedenen Energien gehörige Bahnen in der Phasenebene

Wirkt keine Kraft, dann ist der *Impuls* $\boldsymbol{p}$ eine Erhaltungsgröße, aus $\boldsymbol{F}(\boldsymbol{r}) = 0$ folgt $\mathrm{d}(m\dot{\boldsymbol{s}})/\mathrm{d}t = 0$,

$\boldsymbol{p} = m\dot{\boldsymbol{s}} = \text{konst.}$

Bei einem System von Massen ist bei fehlender äußerer Kraft $\boldsymbol{F}_a$ der Gesamtimpuls $\boldsymbol{P} = \sum m_i \dot{\boldsymbol{s}}_i = M\dot{\boldsymbol{s}}$ eine Erhaltungsgröße.

Besitzt das Kraftfeld ein Potential $V(\boldsymbol{r})$ – man nennt das Kraftfeld dann auch konservativ –, dann ist die *mechanische Energie W* eine Erhaltungsgröße, aus

$$\boldsymbol{F}(\boldsymbol{r}) = -\nabla V(\boldsymbol{r}) \text{ folgt } \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{m}{2}\dot{\boldsymbol{s}}^2 + V(\boldsymbol{s}(t))\right] = 0,$$

$$W = \frac{m}{2}\dot{\boldsymbol{s}}^2 + V(\boldsymbol{s}) \equiv T + V = \text{konst.}$$

Die mechanische Energie ist die Summe von *kinetischer Energie* (T oder W_{kin}) und *potentieller Energie* (V oder W_{pot}). Die kinetische Energie ist positiv definiert, $W_{\text{kin}} \geqq 0$. Die potentielle Energie – und damit auch die Gesamtenergie – enthält eine beliebig wählbare Integrationskonstante. Physikalische Bedeutung hat nur die Änderung der potentiellen Energie bei einer Lageänderung. Wird z. B. die potentielle Energie im Schwerefeld mit $W_{\text{pot}} = mgh$ angegeben, dann wird dem Punkt $h = 0$ der Wert $W_{\text{pot}} = 0$ zugerechnet. Als $h = 0$ wird aber je nach Beispiel der Erdboden, der Experimentiertisch oder ein anderer zweckmäßiger Bezugspunkt gewählt. Man kann dies in der potentiellen Energie zum Ausdruck bringen, indem man $W_{\text{pot}} = mg(h - h_0)$ mit beliebig wählbarem h_0 ansetzt. Legt man den Koordinatenursprung aber in die Höhe h_0, dann erhält man die einfache Formel $W_{\text{pot}} = mgh$ zurück.

Der angegebene Fall, Existenz eines Potentials $V(\boldsymbol{r})$, liegt bei den meisten Beispielen vor. Die mechanische Energie ist aber unter weniger einschneidenden Bedingungen Erhaltungsgröße. Für die Energiebilanz spielt nur die Kraftkomponente tangential zur Bahnkurve F_t eine Rolle. Nur sie leistet wegen $\dot{\boldsymbol{s}} \cdot \boldsymbol{F} = \dot{\boldsymbol{s}} \cdot \boldsymbol{F}_t$ Arbeit und trägt zur Änderung der kinetischen Energie bei $\frac{\mathrm{d}}{\mathrm{d}t}\frac{m}{2}\dot{\boldsymbol{s}}^2 = \dot{\boldsymbol{s}} \cdot \boldsymbol{F} = \dot{\boldsymbol{s}} \cdot (\boldsymbol{F}_{\text{kons.}} + \boldsymbol{F}_{\text{nichtkons.}})_t$. So geht die Kraft auf eine Ladung im Magnetfeld nicht in die Energiebilanz ein.

Damit die Energie eine Erhaltungsgröße ist, darf die Tangentialkomponente der Kraft $\boldsymbol{F}_t$ nicht von der Zeit oder der Geschwindigkeit abhängen, und sie muß ein Potential aufweisen. Tritt in $\boldsymbol{F}_t$ ein nichtkonservativer Anteil $\boldsymbol{F}_{\text{nichtkons.}}$ auf, z. B. eine Reibungskraft, dann ist die mechanische Energie keine Erhaltungsgröße. Beispielsweise wird bei der Reibung mechanische Energie in Wärmeenergie umgewandelt.

In einem Zentralkraftfeld ist der *Drehimpuls* $\boldsymbol{L}$ eine Erhaltungsgröße, aus $\boldsymbol{F}(\boldsymbol{r}) = F(r)\boldsymbol{e}_r$ folgt $\mathrm{d}(\boldsymbol{s} + \boldsymbol{p})/\mathrm{d}t = 0$,

$\boldsymbol{L} = \boldsymbol{s} \times \boldsymbol{p} = \text{konst.}$

Da die genannten Fälle bei den auftretenden Kraftfeldern häufig realisiert sind, sind Energie, Impuls und Drehimpuls wichtige physikalische Größen.

2.1.3.8. Phasenebene. Bei einer eindimensionalen Bewegung im Potential $V(x)$ lassen sich durch Betrachtung der Phasenebene, d. h. der Impuls-Ortsebene, weitgehende Aussagen über die möglichen Bahnkurven gewinnen (Abb. 2.1.3–15b). Man löst die Energie $W(p, x)$ nach p auf, $p = \pm\sqrt{2m[W - V(x)]}$. Geschlossenen Kurven entsprechen *gebundene Zustände*. Kann die Masse bis ins Unendliche fliegen, dann spricht man von *Streuzuständen*. In Abb. 2.1.3–15 existieren für $W < V_0$ keine Bewegungszustände. Für $V_0 \leqq W < 0$ gibt es nur gebundene Zustände. Für $0 \leqq W < V_1$ treten neben gebundenen Zuständen auch Streuzustände auf. Für $W < V_2$ wird ein einlaufendes Teilchen zurückgestreut, für $W > V_2$ kann es durchlaufen.

Bei gebundenen Zuständen ist aus der umschlossenen Fläche in der Phasenebene $I(W) = \oint \mathrm{d}x p(x)$, die Dauer $T = \mathrm{d}I(W)/\mathrm{d}W$ für eine Bewegungsperiode berechenbar.

2.1.4. Bewegung im konstanten Kraftfeld

In einem konstanten Kraftfeld ist die Kraft auf eine Masse unabhängig von deren Position. Die zu berechnende Funktion $\boldsymbol{s}(t)$ ist dann nur als Beschleunigung im Newtonschen Grundgesetz enthalten. Eine direkte Integration ist möglich.

2.1.4.1. Kräftefreie Bewegung. Für $\boldsymbol{F}(\boldsymbol{r}) = 0$ hat die Bewegungsgleichung die einfache Form $m\ddot{\boldsymbol{s}} = 0$. *Der Impuls $\boldsymbol{p} = m\dot{\boldsymbol{s}}$ ist eine Erhaltungsgröße* (1. Newtonsches Axiom). Die Bahnkurven sind Geraden, $\boldsymbol{s}(t) = (\boldsymbol{p}/m)t + \boldsymbol{s}_0 = \boldsymbol{v}t + \boldsymbol{s}_0$, die mit konstanter Geschwindigkeit $\boldsymbol{v} = \boldsymbol{p}/m$ durchlaufen werden. Bei der Betrachtung der Bewegung einer Masse arbeitet man daher einfacher mit der Geschwindigkeit statt mit dem Impuls. Die mechanische Energie $W = p^2/2m = mv^2/2$ ist ebenfalls eine Erhaltungsgröße. Sie ist hier durch die kinetische Energie $p^2/2m$ allein gegeben.

2.1.4.2. Senkrechter Fall und Wurf. Eine Masse, die entlang einer Schiene reibungsfrei senkrecht fallen oder steigen kann, hat einen Freiheitsgrad $z(t)$. Auf die Masse wirkt die Schwerkraft $m\boldsymbol{g} = mg\boldsymbol{e}_z$. Sie hat in diesem Kraftfeld die *potentielle Energie* $W_{\text{pot}}(z) = mgz$. Das Potential wurde also so gewählt, daß die potentielle Energie für $z = 0$ den Wert Null hat. Die Differenz der Potentialwerte für verschiedene Höhen $z_2 - z_1 = h$ entspricht der *Hubarbeit mgh*.

Die *mechanische Energie ist eine Erhaltungsgröße*. Sind zu einem Zeitpunkt Lage z und Geschwindigkeit $\dot{z}$ bekannt, dann kann der Wert von W angegeben werden.

Freier Fall. Beim freien Fall aus der Höhe $z = h$ mit der Anfangsgeschwindigkeit $\dot{z} = 0$ hat die Energie den Wert $W(0, h) = mgh$. Zu Beginn des Falles ist sie gleich der potentiellen Energie. Beim Auftreffen bei $z = 0$ ist $\dot{z} < 0$. Die Energie ist durch $W(z, 0) = m\dot{z}^2/2$ gegeben, sie ist gleich der kinetischen Energie. Der Betrag der Geschwindigkeit $\dot{z}(z = 0) = v$ ergibt sich aus der Erhaltung der Energie. Mit dem bekannten Wert von W folgt $mgh = mv^2/2$ und damit die *Fallgeschwindigkeit* $v = \sqrt{2gh}$.

Wurf. Bei einem Wurf von $z = 0$ mit $\dot{z} = v > 0$ beginnend ist die Energie am Anfang des Wurfes gleich der kinetischen Energie $W(v, 0) = W_{kin} = mv^2/2$. Beim Steigen nimmt die Geschwindigkeit ab, die kinetische Energie wird kleiner; dafür wächst die potentielle Energie an. Im höchsten Punkt ist die Geschwindigkeit auf Null abgesunken, die kinetische Energie ist vollständig in potentielle Energie umgewandelt worden, $W(0, h) = W_{pot} = mgh$. Daraus folgt die *Steighöhe* $h = v^2/2g$. Ein Stein, der mit einer Geschwindigkeit von 10 m/s nach oben geworfen wird, erreicht also eine Höhe von ≈ 5 m.

Bereits aus diesen wenigen Beispielen ist ersichtlich, daß allein durch Verwendung von Erhaltungsgrößen schon viele Aussagen gewonnen werden können. Will man den Zeitablauf der Bewegung betrachten, dann muß man die *Bahnkurve* selbst berechnen. Die Lösung der Bewegungsgleichung für den senkrechten Fall und Wurf ist $z(t) = -gt^2/2 + vt + z_0$, z_0 und v bedeuten hierin Höhe und Geschwindigkeit zur Zeit $t = 0$. Für den oben betrachteten freien Fall gilt also $z(t) = -gt^2/2 + h$. Die Stelle $z = 0$ wird daher nach der Fallzeit $t_F = \sqrt{2h/g}$ erreicht. Für den oben betrachteten Wurf gilt $z(t) = -gt^2/2 + vt$ bzw. $\dot{z}(t) = -gt + v$. Die größte Höhe – durch $\dot{z} = 0$ gekennzeichnet – ist nach der *Steigzeit* $t_S = v/g$ erreicht (Abb. 2.1.4–1). Sie beträgt $h = z(t_S) = gt_S^2/2 = v^2/2g$. Bei gleicher Höhe ist die Steigzeit also gleich der Fallzeit.

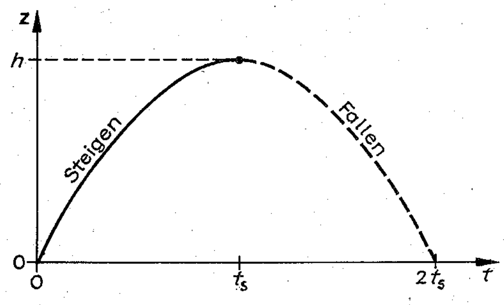


Abb. 2.1.4–1 Bahnkurve eines Wurfes

2.1.4.3. Schiefe Ebene. Der von einer schiefen Ebene auf einer Schiene abgleitende Körper (vgl. Abb. 2.1.3–7) hat einen Freiheitsgrad $s(t)$. Sieht man von der Reibung ab, dann lautet die Bewegungsgleichung $M\ddot{s} = Mg\sin\alpha$. Nur die Hangabtriebskraft beschleunigt den Körper. Die Energie ist eine Erhaltungsgröße $W(\dot{s}, s) = M\dot{s}^2/2 + Mgs\sin\alpha = \text{konst.}$ Die potentielle Energie ist durch das Potential der Schwerkraft gegeben, da $s\sin\alpha$ die Höhe der Masse bedeutet.

Über den Bewegungsablauf lassen sich alle in Abschnitt 2.1.4.2. gemachten Aussagen wiederholen, man braucht nur in allen Ergebnissen g durch $g\sin\alpha$ zu ersetzen.

2.1.4.4. Räumlicher Wurf. Bei einem räumlichen Wurf hat eine Masse drei Freiheitsgrade. Die Bewegungsgleichung

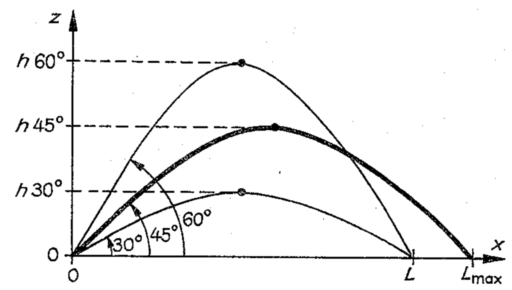


Abb. 2.1.4–2 Wurfparabeln bei gleicher Geschwindigkeit, jedoch verschiedenen Abwurfwinkeln

und ihre Lösung lauten $m\ddot{\boldsymbol{s}} = m\boldsymbol{g}$, $\boldsymbol{s}(t) = \boldsymbol{g}t^2/2 + \boldsymbol{v}t + \boldsymbol{s}_0$. Die Bahnkurve enthält Anfangsort $\boldsymbol{s}_0$ und Anfangsgeschwindigkeit $\boldsymbol{v}$, das sind die Werte von $\boldsymbol{s}(t)$ bzw. $\dot{\boldsymbol{s}}(t)$ für $t = 0$. Legt man die z-Achse in die Vertikale, dann ist $\boldsymbol{g} = -g\boldsymbol{e}_z$, d. h. $F_x = F_y = 0$ verschwinden. Daher sind die Impulskomponenten p_x und p_y Erhaltungsgrößen. In der Draufsicht erhält man in der x-y-Ebene eine geradlinige Bewegung mit konstanter Geschwindigkeit. Legt man die x-Achse in diese Gerade, womit also $v_y = 0$ und $y_0 = 0$ werden, dann erfolgt die Bewegung in der x-z-Ebene. $z(t) = -gt^2/2 + v_z t + z_0$, $x(t) = v_x t + x_0$ sind die Komponenten der Bahnkurve. Diese beiden Gleichungen sind eine Parameterdarstellung einer Parabel (Abb. 2.1.4–2). In z-Richtung hat man eine Bewegung wie beim freien Fall. Sie wird überlagert durch eine Bewegung mit konstanter Geschwindigkeit in x-Richtung. Die *Wurfhöhe* $h = v_z^2/2g$ wird durch v_z bestimmt, die *Wurfweite* $L = x(2t_S)$ durch $L = 2v_x v_z/g = v^2\sin(2\alpha)/g$. Bei gegebener Anfangsgeschwindigkeit ist die Wurfweite für $\alpha = 45°$ am größten, und es kann eine Weite L mit zwei Abwurfwinkeln erreicht werden (vgl. Abb. 2.1.4–2). Die Luftreibung verändert aber diese Ergebnisse.

2.1.4.5. Reibung. Müssen Reibungskräfte einbezogen werden, dann ist die mechanische Energie keine Erhaltungsgröße mehr. Reibungskräfte haben kein Potential, sie führen nicht zur Umwandlung von kinetischer Energie in potentielle Energie, sondern zur Erzeugung von Wärmeenergie. Um die entsprechende Energiemenge nimmt die mechanische Energie ab.

Beim Abgleiten auf der *schiefen Ebene* ist die Reibungskraft $F_R = -\mu Mg\cos\alpha$. Sie ist der Geschwindigkeit entgegengerichtet, d. h. geschwindigkeitsabhängig. Die Bewegungsgleichung $M\ddot{s} = Mg\sin\alpha - \mu Mg\cos\alpha$, $\ddot{s} = g\sin\alpha(1 - \mu\cot\alpha)$ drückt aus, daß weiterhin eine gleichförmig beschleunigte Bewegung vorliegt, die Beschleunigung aber kleiner als ohne Reibung ist. Wird eine Masse eine schiefe Ebene hinauf gezogen (Abb. 2.1.4–3), dann wird die Bewegung durch

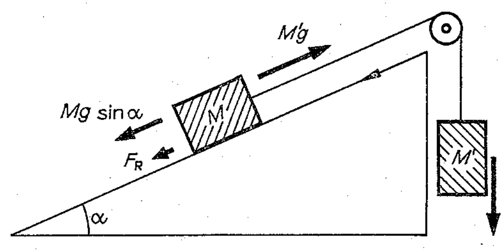


Abb. 2.1.4–3 Aufwärtsbewegung auf einer schiefen Ebene mit Reibung

$(M + M')\ddot{s} = Mg\sin\alpha + \mu Mg\cos\alpha - M'g$ bestimmt. Von der Masse der Rolle soll abgesehen werden können. Damit sich die Masse M bergauf bewegt, muß $M' > M(\sin\alpha + \mu\cos\alpha)$ gelten.

Beim senkrechten Fall oder Wurf mit *geschwindigkeitsproportionaler Luftreibung* $F_R = -\gamma m\dot{z}$ lautet die Bewegungsgleichung $m\ddot{z} = -mg - \gamma m\dot{z}$. Die mechanische Energie nimmt im Zeitablauf ab, $dW/dt = -\gamma 2W_{kin}$, wobei die Verlustleistung proportional zur kinetischen Energie ist.

Die Bahnkurve hat eine mathematisch komplizierte Form

$$z = -\frac{v + g/\gamma}{\gamma} e^{-\gamma t} - \frac{g}{\gamma} t + \frac{v + g/\gamma}{\gamma} + z_0,$$

$$t_S = \frac{1}{\gamma} \ln\left(1 + \frac{\gamma v}{g}\right), \quad h = \frac{v}{\gamma} - \frac{g}{\gamma^2} \ln\left(1 + \frac{\gamma v}{g}\right).$$

Steigzeit und Wurfhöhe haben eine wesentlich andere Form als beim reibungsfreien Wurf. Für den Fall mit geschwindigkeitsproportionaler Reibung ist charakteristisch, daß sich für Zeiten $t \gg 1/\gamma$ ein Fall mit konstanter Geschwindigkeit $\dot{z} = -g/\gamma$ einstellt. Die Zeitabhängigkeit der Bewegungen Steigen und Fallen unterscheidet sich voneinander.

2.1.4.6. Raketen. Raketen werden durch den Rückstoß ausgestoßener Gase angetrieben. Die Masse der Rakete, deren Bahnkurve allein interessiert, ist also zeitabhängig, sie nimmt im Zeitablauf ab. Bei der Betrachtung wird davon ausgegangen, daß der Massenausstoß zeitlich konstant ist, $dm/dt = -\mu =$ konst. bzw. $m(t) = m_0 - \mu t$. Auch die Ausstoßgeschwindigkeit relativ zur Rakete soll einen konstanten Wert $q =$ konst. aufweisen. Für einen außenstehenden Beobachter haben die Gase eine Geschwindigkeit $v - q$, wenn v die Geschwindigkeit der Rakete ist.

In einem Zeitintervall dt stößt die Rakete die Masse $\mu dt = m(t) - m(t + dt)$ aus. Der Impuls der zum Zeitpunkt t betrachteten Masse $m(t)$ ist daher zum Zeitpunkt $t + dt$ durch $p(t + dt) = m(t + dt)v(t + dt) + \mu dt(v - q) = p(t) + m\dot{v}dt - \mu q dt$ gegeben. Die Impulsänderung $[p(t + dt) - p(t)]/dt$ ist nach der Newtonschen Gleichung gleich der äußeren Kraft F. Daher folgt $m\dot{v} = F + \mu q$.

Als antreibenden Rückstoß ergibt diese Betrachtung μq.

Die Diskussion soll auf die Bewegung in Höhen beschränkt werden, in denen man $F = -mg$ setzen kann. Damit die Rakete überhaupt aufsteigt, muß $\mu q > m_0 g$ gelten. Für die Geschwindigkeit folgt aus der Bewegungsgleichung mit $v(0) = 0$, $v(t) = -gt - q\ln(1 - \mu t/m_0)$, für die Bahnkurve mit $z(0) = 0$,

$z(t) = qt + (m_0 - \mu t)(q/\mu)\ln(1 - \mu t/m_0) - gt^2/2$.

Zu dem Zeitpunkt, zu dem das Triebwerk abgestellt wird, muß man die Lösung für den senkrechten Wurf anschließen. Will man ein eventuelles Abbremsen beschreiben, muß man bei der Lösung nur andere Anfangsbedingungen ansetzen.

2.1.5. Bewegung unter dem Einfluß einer Federkraft

In einer stabilen Gleichgewichtslage (vgl. Abb. 2.1.3-15) hat die potentielle Energie ein Minimum. Bei einer kleinen Auslenkung aus der Gleichgewichtslage treten rücktreibende Kräfte auf, die die Masse zur Gleichgewichtslage x_0 zurückziehen. In der Umgebung der Gleichgewichtslage kann man die Potentialfunktion $V(x)$ durch eine Parabel annähern, wie aus der Taylor-Entwicklung $V(x) = V(x_0) + (x - x_0)^2 V''(x_0)/2 + \dots$ erkennbar ist. Bei kleinen Auslenkungen aus der Gleichgewichtslage treten also Bewegungen auf, die durch das Kraftfeld $F = -kx$, $V(x) = (k/2)x^2$ mit der *Federkonstanten* $k > 0$ charakterisiert sind. Da diese Überlegungen für alle Bewegungen um Gleichgewichtslagen gelten, sind die zugehörigen Bewegungsformen, die *harmonischen Schwingungen*, weit verbreitet. Bei einer Bewegung im Dreidimensionalen ist die Potentialfunktion in der Umgebung der Gleichgewichtslage durch

$V(\boldsymbol{r}) = \sum_{i,j} \frac{1}{2} x_i K_{ij} x_j = \frac{1}{2} \boldsymbol{r}K\boldsymbol{r}$ gegeben. Der symmetrische *Kraftkonstantentensor* ist durch die Richtung seiner drei Hauptachsen und seine drei Eigenfrequenzen charakterisiert. Bei einer räumlichen Bewegung überlagern sich Schwingungen zu diesen drei Frequenzen.

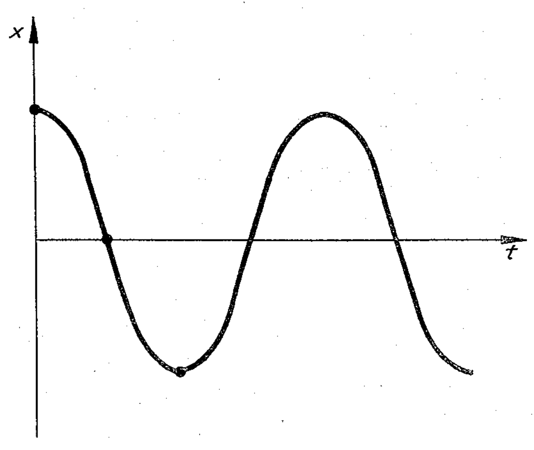


Abb. 2.1.5-1 Harmonischer Oszillator

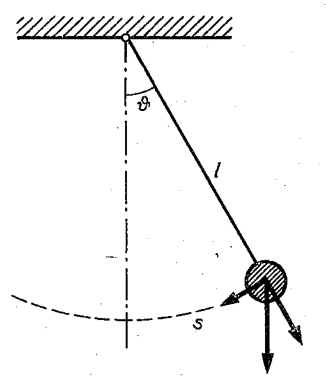


Abb. 2.1.5-2 Ebenes Pendel

2.1.5.1. Harmonische Schwingung. Die Bewegung des *harmonischen Oszillators*, d. h. die Bewegung unter dem Einfluß einer Federkraft $F = -kx$, steht stellvertretend für alle Bewegungen um Gleichgewichtslagen. Unter Einführung einer *Kreisfrequenz* ω läßt sich die Bewegungsgleichung $m\ddot{x} = -kx$ für diese Bewegung mit einem Freiheitsgrad in $\ddot{x} + \omega^2 x = 0$, $\omega = \sqrt{k/m}$, vereinfachen. Die Lösung $x(t) = a\cos(\omega t - \varphi)$ enthält mit der *Amplitude a* und der Phase φ zwei Integrationskonstanten, die durch die speziellen Anfangsbedingungen bestimmt sind. Abb. 2.1.5-1 zeigt für $\varphi = 0$ die Bahnkurve und die Zustände für $\omega t = 0, \pi/2, \pi$.

Wirken *Federkraft* und *Schwerkraft* gleichzeitig, $F = -kx - mg$, dann verschiebt sich nur der Punkt, um den die

Masse schwingt, da $F = -k(x - x_0)$ mit $x_0 = -mg/k$ schreibbar ist.

Ein *ebenes Pendel* hat einen Freiheitsgrad, die Lage ist durch den Winkel ϑ (Abb. 2.1.5–2) bzw. die Bogenlänge $s = l\vartheta$ bestimmt. Die Kraftkomponente in Bewegungsrichtung ist $F = -mg\sin\vartheta$. Bei kleinen Auslenkungen geht die Bewegungsgleichung $m\ddot{s} = F = -mg\vartheta = -m(g/l)s$ in die Schwingungsgleichung mit der Kreisfrequenz

$$\omega = \sqrt{g/l},\quad T = 2\pi\sqrt{l/g}$$

über. Aus der Schwingungsdauer T eines Pendels läßt sich bei bekannter Fadenlänge l die Schwerebeschleunigung g gewinnen.

Bei der Bewegung unter dem Einfluß einer Federkraft ist die *mechanische Energie eine Erhaltungsgröße* $W = (m/2)\dot{x}^2 + (k/2)x^2 = (m/2)\omega^2 a^2$. In dem in Abb. 2.1.5–1 gezeigten Beispiel ist zu den Zeitpunkten $\omega t = 0, \pi$ nur potentielle Energie vorhanden, zum Zeitpunkt $\omega t = \pi/2$ nur kinetische Energie. Im Zeitmittel über eine Schwingungsperiode T sind potentielle und kinetische Energie gleich groß.

2.1.5.2. Gedämpfte harmonische Schwingung. Die Erfahrung lehrt, daß jede Schwingung abklingt, wenn sie nicht durch ständige Energiezufuhr aufrecht erhalten wird. Sie ist gedämpft.

Die *Reibungskraft* erweist sich in vielen Fällen zur Geschwindigkeit proportional $F_R = -m\gamma\dot{x}$, so daß die Bewegungsgleichung für den *gedämpften harmonischen Oszillator* $m\ddot{x} + m\gamma\dot{x} + m\omega_0^2 x = 0$ lautet. Die Lösung für $\omega_0 > \gamma/2$, $x(t) = a\,e^{-(\gamma/2)t}\cos(\omega t - \varphi)$, $\omega = \sqrt{\omega_0^2 - (\gamma/2)^2}$ ist in Abb. 2.1.5–3 dargestellt. Es ist eine Schwingung mit der Kreisfrequenz ω, deren Amplitude exponentiell abklingt. $x(t) = a_+ e^{-\lambda_+ t} + a_- e^{-\lambda_- t}$, $\lambda_\pm = \gamma/2 \pm \sqrt{(\gamma/2)^2 - \omega_0^2}$ erhält man für $\omega_0 < \gamma/2$, im aperiodischen Grenzfall $\omega_0 = \gamma/2$

$$x(t) = \left[x_0\left(1 + \frac{\gamma}{2}t\right) + v_0 t\right] e^{-\gamma t/2}.$$

Bei *starker Dämpfung* bildet sich keine Schwingung mehr aus, es gibt höchstens einen Nulldurchgang.

Die Konstanten a, φ bzw. a_+, a_- bzw. x_0, v_0 werden durch die Anfangsbedingungen festgelegt. Die *mechanische Energie* $W = (m/2)\dot{x}^2 + (m/2)\omega_0^2 x^2$ nimmt bei der gedämpften Bewegung im Zeitablauf ab $\mathrm{d}W/\mathrm{d}t = -2\gamma W_{kin} < 0$.

Bei *schwacher Dämpfung* $\gamma/2 \ll \omega_0$, d. h. wenn sich zwei aufeinanderfolgende Amplituden nur geringfügig unterscheiden, ist $W(t) = (m/2)\omega_0^2 a^2 \exp(-\gamma t)$; in der Zeit $\tau = 1/\gamma$ fällt die mechanische Energie auf den e-ten Teil ab.

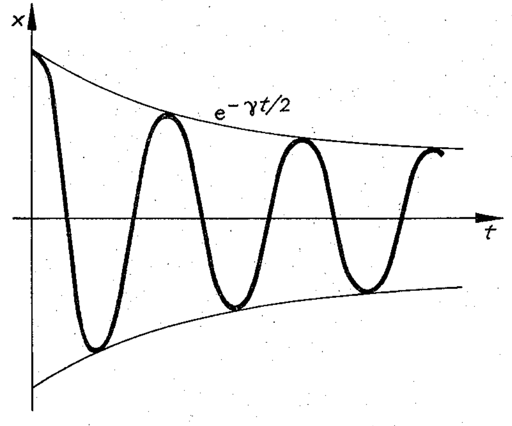
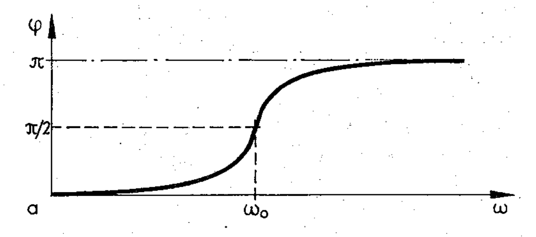
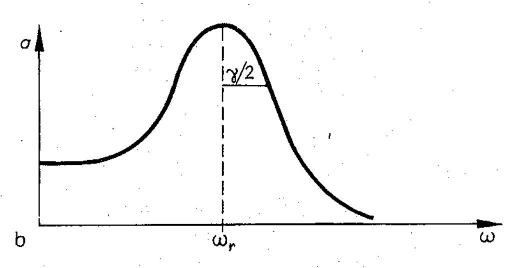


Abb. 2.1.5–3 Gedämpfte harmonische Bewegung





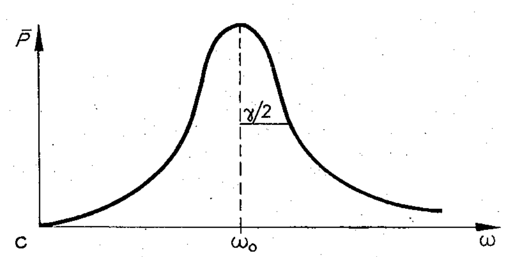


Abb. 2.1.5–4 (a) Phase, (b) Amplitude und (c) absorbierte Energie bei einer erzwungenen Schwingung

Die das schwingende System charakterisierenden *Materialkonstanten* ω_0 (bzw. k) und γ kann man durch Messung der Schwingungsdauer und des Verhältnisses der Amplituden zweier aufeinanderfolgender Ausschläge bestimmen.

2.1.5.3. Erzwungene Schwingungen. Wirkt auf einen gedämpften harmonischen Oszillator eine *periodische* Kraft, dann erhält man die Lösung der Bewegungsgleichung

$$\left(\frac{\mathrm{d}^2}{\mathrm{d}t^2} + \gamma\frac{\mathrm{d}}{\mathrm{d}t} + \omega_0^2\right)x(t) = \frac{F_0}{m}\cos\omega t,$$

indem man zu der Lösung der homogenen Gleichung (vgl. 2.1.5.2.) noch eine spezielle Lösung der inhomogenen Gleichung hinzufügt: $x(t) = x_{hom} + x_{inhom}$. Für Zeiten $\gamma t \gg 1$ ist die durch die speziellen Anfangsbedingungen geprägte homogene Lösung abgeklungen, es wird $x = x_{inh}$. Diese Lösung lautet $x_{inh}(t) = a\cos(\omega t - \varphi)$.

Es ist eine erzwungene Schwingung mit der Frequenz ω der äußeren Kraft. Amplitude a und Phase φ hängen von dieser Frequenz ab,

$$a(\omega) = \frac{F_0/m}{[(\omega_0^2 - \omega^2)^2 + (\gamma\omega)^2]^{1/2}},$$

$$\tan\varphi(\omega) = \frac{\gamma\omega}{\omega_0^2 - \omega^2}$$

Abb. 2.1.5–4b zeigt, daß die Amplitude in einem durch die Dämpfung bestimmten Bereich um die Eigenfrequenz ω_0

besonders groß ist (Resonanz). Die Phasenverschiebung zur Kraft (Abb. 2.1.5–4a) hat bei $\omega = \omega_0$ den Wert $\pi/2$.
Um die *Eigenfrequenz* und die *Dämpfung* zu ermitteln (die Näherungsausdrücke beziehen sich immer auf $\gamma \ll \omega_0$), kann man die *Resonanzfrequenz* $\omega_r = \sqrt{\omega_0^2 - \gamma^2/2} \approx \omega_0$, bei der die Amplitude am größten ist, und die *Resonanzüberhöhung* $a(\omega_r)/a(0) = (\omega_0/\gamma)\omega_0/[\omega_0^2 - (\gamma/2)^2]^{1/2} \approx \omega_0/\gamma$ oder die *Resonanzbreite* $a(\omega_r \pm \gamma/2) \approx a(\omega_r)/\sqrt{2}$ messen. Vielfach wird aber auch die *absorbierte Energie* bestimmt.
Die Änderung der mechanischen Energie im Zeitablauf $\mathrm{d}W/\mathrm{d}t = -m\gamma\dot{x}^2 + \dot{x}F$ wird durch die – durch die Reibung bedingte – Verlustleistung $P_R = m\gamma\dot{x}^2 = ma^2\omega^2\gamma\sin^2(\omega t - \varphi) \geqq 0$ und durch die aus dem äußeren Kraftfeld absorbierte Leistung $P_{abs} = \dot{x}F = ma^2\omega^2\gamma\sin^2(\omega t - \varphi) - (m/2)a^2\omega(\omega_0^2 - \omega^2)\sin[2(\omega t - \varphi)]$ bestimmt. In P_{abs} ist der erste Summand positiv und gleich der Verlustleistung. Der zweite Summand beschreibt einen Energieaustausch zwischen Oszillator und Kraftfeld. Er verschwindet im Zeitmittel. Die im Zeitmittel absorbierte Energie $\bar{P}_{abs}(\omega) = (m/2)\gamma a^2(\omega)\omega^2$ ist im Resonanzfall $\omega = \omega_0$ am größten, bei $\omega = \sqrt{\omega_0^2 + (\gamma/2)^2} \pm \gamma/2 \approx \omega_0 \pm \gamma/2$ ist sie auf den halben Wert abgesunken (Abb. 2.5.1–4c).

2.1.5.4. Gekoppelte Oszillatoren. Abb. 2.1.5–5 zeigt zwei Oszillatoren, die über eine Feder mit der Federkonstanten

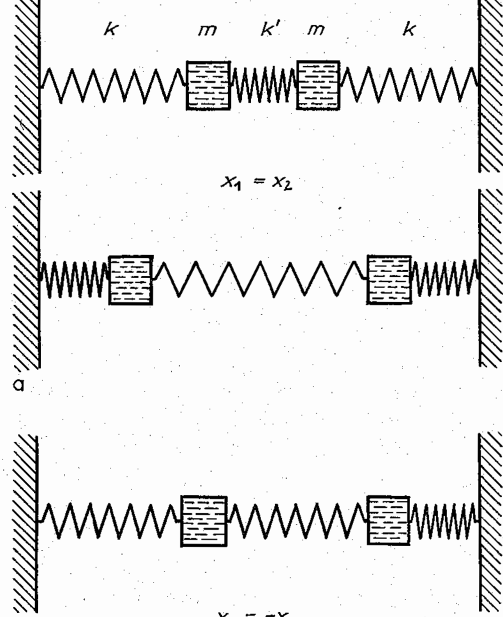


Abb. 2.1.5–6 (a) symmetrische und (b) antisymmetrische Schwingung

k' gekoppelt sind. Für den Fall gleicher Massen $m_1 = m_2 = m$ und gleicher Federkonstanten $k_1 = k_2 = k$ läßt sich anschaulich erkennen, daß die in Abb. 2.1.5–6 dargestellte *symmetrische* und *antisymmetrische Schwingung* Lösungen der Bewegungsgleichung sind. Im Fall (b) wird die Feder k' überhaupt nicht beansprucht, die Massen schwingen dann

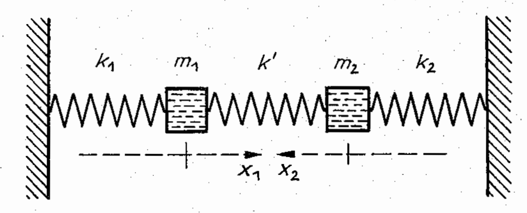


Abb. 2.1.5–5 Gekoppelte Oszillatoren

mit der ohne Kopplung auftretenden Eigenfrequenz $\omega_- = \sqrt{k/m}$. Im Fall (a) verläuft die Bewegung in gleicher Weise, wenn die Feder k' in der Mitte fest eingespannt wird. Da die halb so lange Feder die doppelte Federkonstante $2k'$ hat, führt die Masse m eine Schwingung aus wie unter dem Einfluß zweier Federn k und $2k'$, deren Kräfte sich addieren. Die Eigenfrequenz der symmetrischen Schwingung ist also $\omega_+ = \sqrt{(k + 2k')/m}$. Sie liegt höher als ω_-.
Eine beliebige Schwingung ist eine Linearkombination der beiden Eigenschwingungen $x_1(t) = A\cos(\omega_+ t - \varphi_+) + B\cos(\omega_- t - \varphi_-)$, $x_2(t) = A\cos(\omega_+ t - \varphi_+) - B\cos(\omega_- t - \varphi_-)$. Bei den Anfangsbedingungen $x_1(0) = 0$, $x_2(0) = 0$, $\dot{x}_1(0) = v$ und $\dot{x}_2(0) = 0$ ist zum Anfangszeitpunkt keine Masse ausgelenkt, der Masse Eins wurde die Geschwindigkeit v erteilt. Es entsteht die in Abb. 2.1.5–7 dargestellte *Schwebung*, zu der die Lösungen $x_1(t) = (v/2\omega_+)\sin\omega_+ t + (v/2\omega_-)\sin\omega_- t$, $x_2(t) = (v/2\omega_+)\sin\omega_+ t - (v/2\omega_-)\sin\omega_- t$ gehören. Bei schwacher Kopplung $k' \ll k$ gilt mit $\omega_\pm = \bar{\omega} \pm \Delta\omega$ genähert $x_1 = (v/\bar{\omega})\sin\bar{\omega}t\cos\Delta\omega t$, $x_2 = -(v/\bar{\omega})\cos\bar{\omega}t\sin\Delta\omega t$. Die Periode der Schwebung ist durch die Stärke der Kopplung bestimmt, $\Delta\omega/\bar{\omega} = k'/2k$.
Können im Fall gleicher Massen und Federkonstanten die Eigenfrequenzen unter Ausnutzung der Symmetrie sofort angegeben werden, so ist dies im allgemeinen Fall nicht möglich. Man muß die Bewegungsgleichungen $m_1\ddot{x}_1 = k_1x_1 - k'(x_1 + x_2)$, $m_2\ddot{x}_2 = -k_2x_2 - k'(x_1 + x_2)$ lösen. Auf jede Masse wirken die Kräfte zweier Federn, wobei die Stauchung der Kopplungsfeder bei der in Abb. 2.1.5–6 verwendeten Koordinatenwahl durch die Summe der Auslenkungen der beiden Massen gegeben ist. Dieses gekoppelte System von Differentialgleichungen hat auch *Eigenschwingungen* als Lösung, d. h. solche Bewegungsformen, bei denen sich beide Massen periodisch mit gleicher Frequenz bewegen. Die zugehörigen *Eigenfrequenzen* sind mit
$\omega_1^2 = (k_1 + k')/m_1$, $\omega_2^2 = (k_2 + k')/m_2$, $\omega_3^2 = k'/\sqrt{m_1m_2}$,
$\omega_\pm^2 = (\omega_1^2 + \omega_2^2)/2 \pm \sqrt{(\omega_1^2 - \omega_2^2)^2/4 + \omega_3^4}$.
Abb. 2.1.5–8 zeigt diese Eigenfrequenzen als Funktion der Kopplungsstärke. Bei verschwindender Kopplung treten die Eigenfrequenzen auf, die zu den getrennten Oszillatoren gehören. Bei starker Kopplung sind Schwerpunkt- und Relativbewegung praktisch getrennt. ω_- beschreibt die Bewegung der Gesamtmasse $m_1 + m_2$ unter dem Einfluß der Federn k_1 und k_2. Die Eigenfrequenz der Relativbewegung ω_+ wird durch das Verhältnis von k' zur reduzierten Masse μ bestimmt.

Abb. 2.1.5–7 Schwebungen

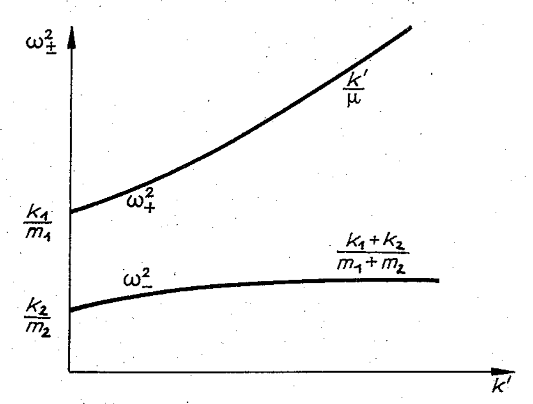


Abb. 2.1.5–8 Eigenfrequenzen bei gekoppelten Schwingungen

2.1.6. Bewegung im Zentralkraftfeld

Zentralkraftfelder, bei denen der Kraftvektor an allen Raumpunkten zu ein und demselben Zentrum weist, treten häufig auf. Als wichtiges Beispiel soll die *Bewegung im Gravitationsfeld*, die Planetenbewegung, betrachtet werden. Von der Mitbewegung der Sonne wird zunächst abgesehen (vgl. 2.1.10.).

2.1.6.1. Keplersche Gesetze. 1. Die Planeten bewegen sich auf Ellipsen, in deren einem Brennpunkt die Sonne steht. 2. Der Leitstrahl von der Sonne zum Planeten überstreicht in gleichen Zeiten gleiche Flächen. 3. Die Quadrate der Umlaufzeiten zweier Planeten verhalten sich wie die dritten Potenzen der großen Halbachsen.

Die in den Keplerschen Gesetzen steckenden Aussagen folgen aus der Newtonschen Gleichung. Newton fand – von den Keplerschen Gesetzen ausgehend – die Abstandsabhängigkeit der Gravitationskraft $F \sim -1/r^2$.

2.1.6.2. Erhaltung des Drehimpulses. Der Drehimpuls $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ ist in einem Zentralkraftfeld eine Erhaltungsgröße (vgl. 2.1.3.7.). Da $\boldsymbol{L}$ ein Vektor ist, enthält die Gleichung $\boldsymbol{L}$ = konst. drei Aussagen: 1. die Richtung von $\boldsymbol{L}$ ist konstant (zwei Aussagen) und 2. der Betrag von $\boldsymbol{L}$ ist konstant (eine Aussage).

Befindet sich eine Masse zur Zeit t am Ort $\boldsymbol{s}(t)$ mit der Geschwindigkeit $\dot{\boldsymbol{s}}(t)$, dann ist der Drehimpuls $\boldsymbol{L} = m\boldsymbol{s} \times \dot{\boldsymbol{s}}$. Er steht senkrecht auf der von den beiden Vektoren $\boldsymbol{s}$ und $\dot{\boldsymbol{s}}$ aufgespannten Ebene. Da sich aber die *Richtung* von $\boldsymbol{L}$ im

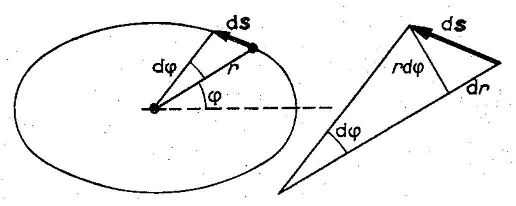


Abb. 2.1.6–1 Wegelement ds und dessen Zerlegung in eine radiale und eine azimutale Komponente

Zeitablauf nicht ändert, liegen s und $\dot{s}$ zu allen Zeiten in der gleichen Ebene. Die Bahnkurve verläuft also in einer Ebene, zu der der Drehimpuls der Normalenvektor ist und in der das Kraftzentrum liegt. Das ist ein Teil der im ersten Keplerschen Gesetz steckenden Aussage.

In dieser Ebene werden Polarkoordinaten eingeführt und das Wegelement ds in einen radialen Anteil dr und einen azimutalen Anteil $r\mathrm{d}\varphi$ zerlegt (Abb. 2.1.6–1). Der Betrag des Drehimpulses hat in diesen Koordinaten den Wert $L = mr^2\dot{\varphi}$ = konst. $r^2\mathrm{d}\varphi/2$ ist die Dreiecksfläche zwischen den zu φ und $\varphi + \mathrm{d}\varphi$ gehörenden Leitstrahlen (Abb. 2.1.6–1), $r^2\dot{\varphi}/2$ also die vom Leitstrahl in dt überstrichene Fläche. Die Aussage, daß der Betrag des Drehimpulses konstant ist, ist also identisch mit dem Inhalt des zweiten Keplerschen Gesetzes.

2.1.6.3. Erhaltungsgrößen im Zentralkraftfeld. Betrachtet werden nur solche Zentralkräfte, die ein Potential aufweisen. Die Flächen konstanten Potentials sind dann Kugeln, d. h., $V = V(r)$. In diesem Kraftfeld sind *Energie* und *Drehimpuls Erhaltungsgrößen* (vgl. 2.1.3.7.). In der Energie $W = p^2/2m + V(r) = (m/2)\dot{s}^2 + V$ wird das Geschwindigkeitsquadrat in Polarkoordinaten ausgedrückt (vgl. Abb. 2.1.6–1) $(\mathrm{d}s/\mathrm{d}t)^2 = \dot{s}^2 = \dot{r}^2 + r^2\dot{\varphi}^2$ und der radiale Impuls $p_r = m\dot{r}$ eingeführt. Aus der Erhaltung von Energie und Betrag des Drehimpulses folgen dann in der Bahnebene

$$W = p_r^2/(2m) + L^2/(2mr^2) + V(r) = \text{konst. und}$$
$$L = mr^2\dot{\varphi} = \text{konst.}$$

$r(t)$ kann aus der ersten Gleichung berechnet werden. Mittels der Erhaltungsgrößen wurde die Aufgabe, ein dreidimensionales Bewegungsproblem zu lösen, auf die Betrachtung der eindimensionalen Bewegung im *Effektivpotential* $V_{\text{eff}}(r) = L^2/(2mr^2) + V(r)$ zurückgeführt. Ist $r(t)$ berechnet, dann kann $\varphi(t)$ aus der zweiten Gleichung gewonnen werden. Man kann auch die Zeit eliminieren. Für die Bahn $r(\varphi)$ erhält man ($r' \equiv \mathrm{d}r/\mathrm{d}\varphi$) als Bestimmungsgleichung $W = L^2/(2mr^4)r'^2 + L^2/(2mr^2) + V(r)$.

2.1.6.4. Bewegung in der Phasenebene. Bei einer eindimensionalen Bewegung $r(t)$ im Effektivpotential $V_{\text{eff}}(r)$ kann man sich die möglichen Bahnkurven am besten in der Phasenebene (vgl. 2.1.3.8.) veranschaulichen. Die Diskussion wird für das Potential $V(r) = -C/r$, $C = GmM$ bzw. $C' = -qQ/4\pi\varepsilon_0$, geführt. Der wesentliche Unterschied zwischen Gravitationsfeld (C) und Coulomb-Kraft (C') besteht darin, daß C stets positiv ist, während C' beide Vorzeichen annehmen kann.

In Abb. 2.1.6–2 ist V_{eff} für verschiedene Drehimpulse dargestellt. Für $L \neq 0$ überwiegt für kleine Abstände das *Zentrifugalpotential* $L^2/2mr^2$, für große r nähert sich V_{eff} der Kurve $-C/r$. Dazwischen durchläuft es ein Minimum bei $r_m = L^2/mC$ und eine Nullstelle $r_0 = L^2/2mC$. Das Minimum $V_{\text{eff}}(r_m) = -C/2r_m$ wird mit wachsendem Drehimpuls immer flacher. Für einen gegebenen Drehimpuls gibt es also keine Lösungen mit $W < W_m = -mC^2/2L^2$, zu einer vorgegebenen Energie $W < 0$ kann der Drehimpuls nur Werte $L^2 \leqq mC^2/2$ annehmen.

Die Darstellung von $p_r(r) = \pm\sqrt{2m[W - L^2/(2mr^2) + C/r]}$ in der Phasenebene (Abb. 2.1.6-3) zeigt, daß es für $W < 0$ gebundene, für $W \geqq 0$ Streuzustände gibt. Die Umkehrpunkte liegen bei $r_\pm = r_m(W_m/W)\left(1 \pm \sqrt{1 - W/W_m}\right)$.

Die zu $W = W_m$ gehörige Lösung $r(t) = r_m$ = konst. besagt, daß in der Bewegungsebene ein Kreis mit dem Radius r_m durchlaufen wird. Aus $L = mr_m^2\,\dot{\varphi}$ und $r_m = L^2/(mC)$ folgt die Bahngeschwindigkeit $v = r_m\dot{\varphi} = \sqrt{C/mr_m}$. Damit also eine

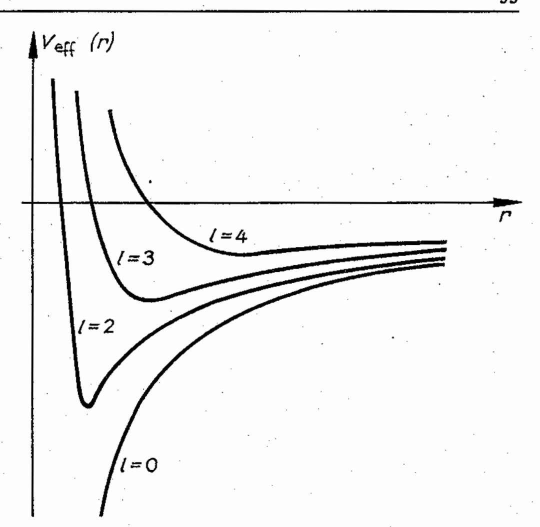


Abb. 2.1.6–2 Effektivpotential zu $V(r) = -C/r$

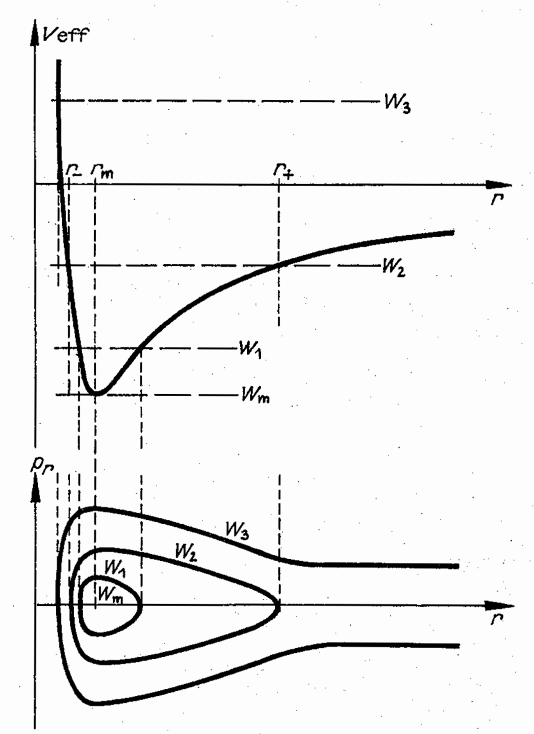


Abb. 2.1.6–3 Gebundene und Streuzustände in der Phasenebene. Für W_2 sind r_- und r_+ eingetragen

Rakete auf einer Kreisbahn um die Erde laufen kann, muß sie die Geschwindigkeit $v = \sqrt{GM_E/R} = \sqrt{gR} = 7{,}9$ km/s (*erste kosmische Geschwindigkeit*) haben.

Für Energien $W > W_m$ liegt die Bahnkurve zwischen zwei Radien r_- und r_+. Im allgemeinen, d. h. für beliebige $V(r)$, ergeben sich Rosettenbahnen (Abb. 2.1.6–4), nur in Spezial-

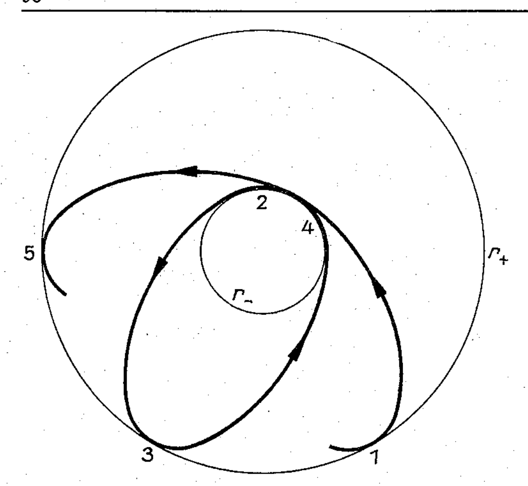


Abb. 2.1.6–4 Rosettenbahn. Die Radien r_+, r_- werden in der Reihenfolge der Numerierung angenommen

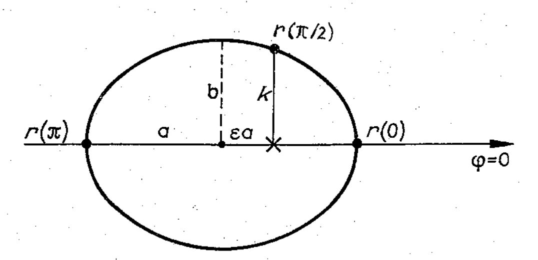


Abb. 2.1.6–5 Ellipsenbahn, Kraftzentrum bei ×

fällen, wie bei $V \sim -1/r$, ergeben sich geschlossene Kurven (Abb. 2.1.6–5). Die Diskussion der Erhaltungsgrößen und der Bewegung in der Phasenebene gestattet es, viele Fragen zu beantworten, ohne die Bahnkurve explizit ausrechnen zu müssen. So hat eine an der Oberfläche der Erde ruhende Masse (die Erddrehung ist hier unwesentlich) die Energie $W = -C/R$. Streuzustände gibt es aber nur für $W \geqq 0$. Daher muß der Masse mindestens die kinetische Energie $W_{\text{kin}} = C/R$ erteilt werden, damit sie das Schwerefeld der Erde verlassen kann. Sie muß mindestens mit der aus $(m/2)v^2 = C/R$ folgenden Geschwindigkeit $v = \sqrt{2gR} = 11{,}2$ km/s (*zweite kosmische Geschwindigkeit*) abgeschossen werden. Die Abschußrichtung spielt dabei keine Rolle (natürlich nicht in die Erde hinein).

2.1.6.5. Planetenbewegung. Die Lösungen der Gleichung für die Bahnkurve $r(\varphi)$ (vgl. 2.1.6.2.) im Gravitationsfeld ($C = GmM$) sind *Kegelschnitte*

$r(\varphi) = k/(1 + \varepsilon\cos\varphi)$, $k = L^2/(mC)$, $k/(\varepsilon^2 - 1) = C/(2W)$.

Die Parameter k und ε sind durch die Werte der Erhaltungsgrößen Energie W und Drehimpuls L (neben den Daten des Systems G, M, m) gegeben.

Für $\varepsilon < 1$ erhält man *Ellipsen*, $r(\varphi)$ bleibt für alle Werte von φ endlich. Der Grenzfall $\varepsilon = 0$ liefert einen *Kreis*. Da $r(\varphi)$ positiv ist, erfordern diese Lösungen $k > 0$, d. h., $C > 0$. Für die Gravitation ist diese Bedingung erfüllt, aber bei gleichnamigen Ladungen gibt es im Coulomb-Feld keine gebundenen Zustände.

Für $\varepsilon \geqq 1$ verschwindet der Nenner für $\cos\varphi = -1/\varepsilon$, die Bahnkurve reicht ins Unendliche. Für $\varepsilon = 1$ erhält man eine *Parabel*, für $\varepsilon > 1$ *Hyperbeln*. Diese Kurven liegen bei Anziehung ($C > 0$, $k > 0$) im Winkelbereich $\cos\varphi > -1/\varepsilon$, bei Abstoßung ($C < 0$, $k < 0$) im Winkelbereich $\cos\varphi < -1/\varepsilon$. Diese Bahnen werden bei der Rutherford-Streuung diskutiert (vgl. 2.1.10.). Gegebene Werte von W und L lassen beide Vorzeichen von ε zu. Da dieses Vorzeichen aber durch eine Transformation $\varphi \to \varphi + \pi$ des Koordinatensystems kompensiert werden kann, ergeben sich keine physikalisch neuen Zustände. Daher beschränkt man sich auf die Betrachtung von $\varepsilon > 0$.

Gebundene Zustände, die Ellipsen, sollen ausführlicher behandelt werden. $\varepsilon < 1$ erfordert $k > 0$, also Anziehung $C > 0$, und wegen $\varepsilon^2 - 1 < 0$ negative Energien $W < 0$. Die große Halbachse der Ellipse (vgl. Abb. 2.1.6-5) $2a = r(0) + r(\pi)$ und der Abstand Brennpunkt – Mittelpunkt $e = a - r(0)$ ergeben sich zu $a = k/(1 - \varepsilon^2)$, $e = \varepsilon a$. Dies macht den Namen *Exzentrizität* für ε verständlich. Der Wert der kleinen Halbachse ergibt sich aus $b^2 = a^2 - e^2$, $b = k/(1 - \varepsilon^2)^{1/2}$. Wegen $r(\pi/2) = k$ kann auch k veranschaulicht werden.

Abb. 2.1.6–6a zeigt *Bahnen gleichen Drehimpulses*, d. h. Lösungen der Bewegungsgleichung zu verschiedenen Energien im gleichen Effektivpotential. Sie sind durch $r(\pi/2) = k =$ konst. gekennzeichnet. Zur minimal möglichen Energie W_m gehört der Kreis.

Abb. 2.1.6–6b zeigt *Bahnen gleicher Energie*, die zu verschiedenen Drehimpulsen gehören. Alle Bahnen haben den gleichen Wert für die große Halbachse a. Diese Mannigfaltigkeit tritt z. B. auf, wenn von einem Punkt aus Massen mit sich nur in der Richtung unterscheidender Geschwindigkeit abgeschossen werden. Für die Energie ist nur der Betrag der Geschwindigkeit maßgebend. Der Drehimpuls wird auch durch deren Richtung bestimmt.

Die Ellipsenfläche ist πab. Mit der Umlaufzeit T lautet daher der Flächensatz $L = 2m(\mathrm{d}A/\mathrm{d}t) = 2m(\pi ab/T)$. Daraus folgt $T^2 = 4\pi^2 a^3/(GM)$. Das ist das dritte Keplersche Gesetz.

2.1.7. Bewegung im elektromagnetischen Feld

Auf eine Ladung wirkt im elektromagnetischen Feld die Lorentz-Kraft $\boldsymbol{F} = q(\boldsymbol{E} + \dot{\boldsymbol{s}} \times \boldsymbol{B})$. In der Bewegungsgleichung $m\ddot{\boldsymbol{s}} = \boldsymbol{F}$ tritt also das Verhältnis q/m, die spezifische Ladung, auf. q/m bzw. m – bei bekannter Ladung – kann also durch Beobachtung der Bewegung im elektromagnetischen Feld gewonnen werden. Dies wird in den Massenspektrografen ausgenutzt. Nachfolgend wird nur die Bewegung im räumlich und zeitlich konstanten Feld diskutiert. Zeitlich veränderliche Felder sind z. B. für die Teilchenbeschleuniger wesentlich.

2.1.7.1. Homogenes Magnetfeld. Da die Kraft $q\dot{\boldsymbol{s}} \times \boldsymbol{B}$ senkrecht zur Bewegungsrichtung steht, $\dot{\boldsymbol{s}} \cdot \boldsymbol{F} = 0$, leistet die magnetische Kraft keine Arbeit, *die Energie ist eine Erhaltungsgröße*. Trotzdem hat diese Kraft kein Potential, wie aus ihrer Geschwindigkeitsabhängigkeit zu ersehen ist.

Die Kraftkomponente in Richtung des Magnetfeldes verschwindet. Legen wir die z-Achse in diese Richtung, $\boldsymbol{B} = B\boldsymbol{e}_z$, dann bewegt sich das Teilchen in z-Richtung kräftefrei mit konstanter Geschwindigkeit. Zu untersuchen bleibt nur die Bewegung in der xy-Ebene. Die Anfangsbedingungen seien $\boldsymbol{s}(0) = 0$ und $\dot{\boldsymbol{s}}(0) = \boldsymbol{v}$, wobei $\boldsymbol{v}$ ein Vektor in der xy-Ebene ist. Die in der Bewegungsgleichung $m\ddot{\boldsymbol{s}} = q\dot{\boldsymbol{s}} \times \boldsymbol{B}$ auftretenden Konstanten lassen sich zu einem

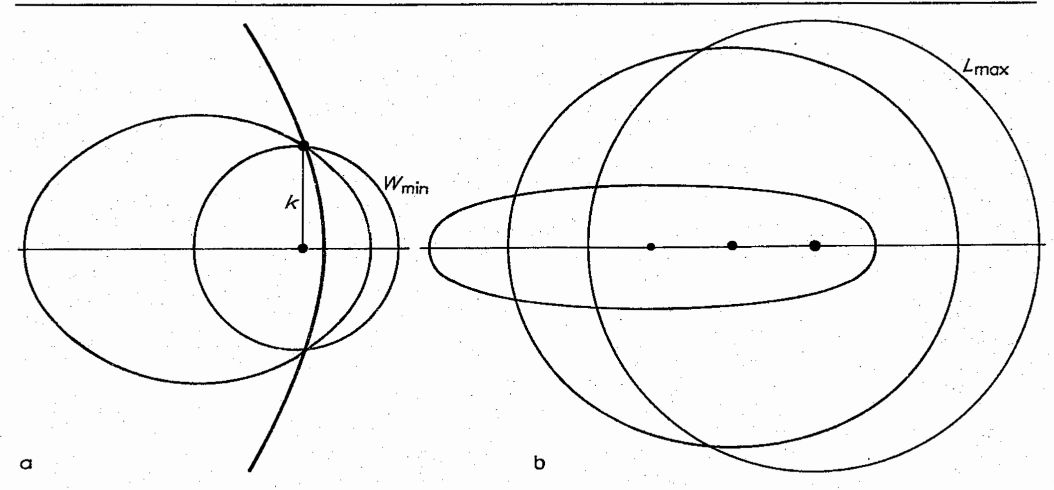


Abb. 2.1.6–6 (a) Bahnen mit gleichem Drehimpuls, W variiert, und (b) Bahnen mit gleicher Energie, L variiert

Vektor
$\boldsymbol{\omega} = -(q/m)\boldsymbol{B}$,
auch *Zyklotronfrequenz* genannt, zusammenfassen. Aus der Energieerhaltung folgt, daß der Betrag der Geschwindigkeit sich im Zeitablauf nicht ändert, $\dot{s}^2 = v^2$. Die Integration der Bewegungsgleichung $\ddot{\boldsymbol{s}} = \boldsymbol{\omega} \times \boldsymbol{s}$ führt auf $\dot{\boldsymbol{s}} = \boldsymbol{\omega} \times (\boldsymbol{s} - \boldsymbol{s}_m)$, $(\boldsymbol{s} - \boldsymbol{s}_m)^2 = R^2 = \text{konst.}, R = v/\omega$.
Wie in Abb. 2.1.7–1 dargestellt, erfolgt die Bewegung auf einem Kreis vom Radius R, dessen Mittelpunkt bei $\boldsymbol{s}_m = (\boldsymbol{\omega} \times \boldsymbol{v})/\omega^2$ liegt. Die Anfangsgeschwindigkeit $\boldsymbol{v}$ ist Tangente an diesen Kreis. Der Kreis wird mit konstanter Geschwindigkeit, d. h. auch mit konstanter Winkelgeschwindigkeit, durchlaufen. Diese Winkelgeschwindigkeit wird durch die Stärke des Magnetfeldes bestimmt. Bei gegebener Feldstärke wächst der Kreisradius mit der Anfangsgeschwindigkeit an. Bei Nebelkammeraufnahmen sind diese Kreise als Spuren geladener Teilchen (vgl. 5.4.4.4.) deutlich auszumachen. Insbesondere ist zu sehen, daß der große Unterschied in den q/m-Werten zu ganz verschiedenen Radien bei Elektronen im Vergleich zu Ionen führt.
Die vollständige Integration der Bewegungsgleichung liefert die Zeitabhängigkeit der Bahnkurve
$\boldsymbol{s}(t) = [(\boldsymbol{\omega} \times \boldsymbol{v})/\omega^2](1 - \cos \omega t) + (\boldsymbol{v}/\omega) \sin \omega t.$
Fügt man eine Bewegung mit konstanter Geschwindigkeit in Richtung des Magnetfeldes hinzu, dann ergibt sich: Im homogenen Magnetfeld durchlaufen geladene Teilchen Spiralbahnen, deren Achse durch die Richtung des Magnetfeldes bestimmt ist. Negativ geladene Teilchen durchlaufen Spiralen mit Rechtsschraubensinn, positiv geladene Teilchen im Linksschraubensinn.

2.1.7.2. Homogenes elektrisches und magnetisches Feld. Aus der Lorentz-Kraft $\boldsymbol{F} = q(\boldsymbol{E} + \dot{\boldsymbol{s}} \times \boldsymbol{B})$ ersieht man, daß die Energie eine Erhaltungsgröße ist, da das zeitlich konstante, homogene elektrische Feld ein Potential hat und das Magnetfeld keine Arbeit leistet. In Richtung von $\boldsymbol{B}$ (z-Richtung) wirkt die Kraftkomponente $F_z = qE_z$, es erfolgt eine gleichmäßig beschleunigte Bewegung. Die Betrachtung kann daher auf den Fall $\boldsymbol{E} \perp \boldsymbol{B} (\boldsymbol{B} = B\boldsymbol{e}_z, \boldsymbol{E} = E\boldsymbol{e}_y)$ beschränkt werden. Mit $\boldsymbol{\omega} = -(q/m)\boldsymbol{B}$ und $\boldsymbol{v}_D = (E/B)\boldsymbol{e}_x = (\boldsymbol{E} \times \boldsymbol{B})/B^2$ ist die Bewegungsgleichung $\ddot{\boldsymbol{s}} = \boldsymbol{\omega} \times (\dot{\boldsymbol{s}} - \boldsymbol{v}_D)$. Bei einer Anfangsgeschwindigkeit $\boldsymbol{v} = \boldsymbol{v}_D$ kompensieren sich dann elektrische und magnetische Kraft, die La-

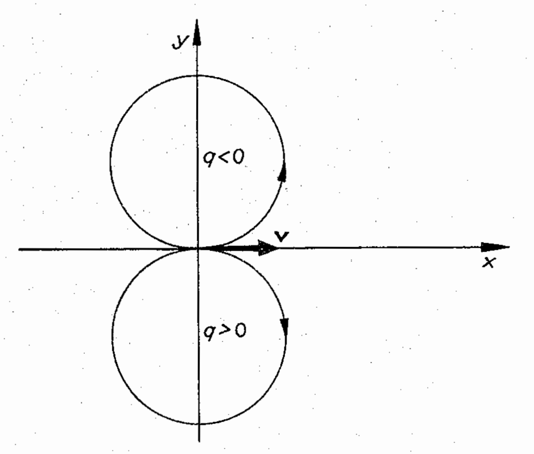


Abb. 2.1.7–1 Zyklotronbewegung im Magnetfeld

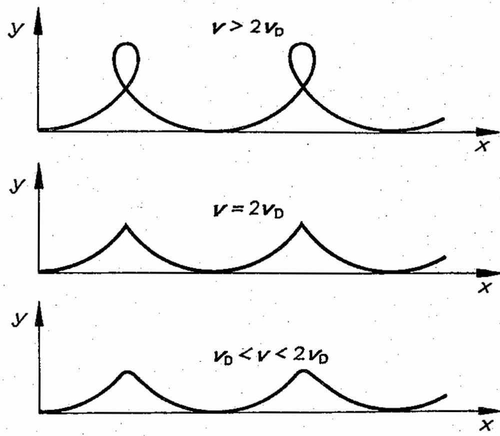


Abb. 2.1.7–2 Zykloidenbahnen

dung bewegt sich kräftefrei mit konstanter Geschwindigkeit. In einem sich mit der *Driftgeschwindigkeit* $\boldsymbol{v}_D$ bewegenden Bezugssystem stimmt die Bewegungsgleichung mit der bei $\boldsymbol{E} = 0$ (vgl. 2.1.7.1.) überein. Die Bahnkurve ist also

$$\boldsymbol{s}(t) = \boldsymbol{v}_D t + \frac{\boldsymbol{\omega} \times (\boldsymbol{v} - \boldsymbol{v}_D)}{\omega^2}(1 - \cos \omega t) + \frac{\boldsymbol{v} - \boldsymbol{v}_D}{\omega} \sin \omega t.$$

Sie ist eine Zykloide, die aus der Überlagerung einer Kreisbewegung zur Anfangsgeschwindigkeit $\boldsymbol{v} - \boldsymbol{v}_D$ und einer Driftbewegung mit der Geschwindigkeit $\boldsymbol{v}_D$ entsteht. Abb. 2.1.7–2 zeigt Beispiele für $\boldsymbol{v} = v\boldsymbol{e}_x$.

2.1.7.3. Larmor-Präzession. Eine Ladung in einem Zentralkraftfeld möge zusätzlich dem Einfluß eines homogenen Magnetfeldes $\boldsymbol{B} = B\boldsymbol{e}_z$ unterworfen werden. Zur Vereinfachung wird die harmonische Kraft $-m\omega_0^2 \boldsymbol{s}$ als Zentralkraft herausgegriffen. Dann lautet die Bewegungsgleichung $m\ddot{\boldsymbol{s}} = -m\omega_0^2 \boldsymbol{s} + q\dot{\boldsymbol{s}} \times \boldsymbol{B}$. Mit welchen Frequenzen ω sind Kreisbewegungen in der xy-Ebene möglich? Nach Einführung der *Larmor-Frequenz* $\boldsymbol{\omega}_L = -(q/2m)\boldsymbol{B}$ lautet die Lösung $\omega = \omega_L \pm \sqrt{\omega_0^2 + \omega_L^2}$. Man erhält dieses Ergebnis, wenn man in die Bewegungsgleichung nach 2.1.2.4. $\ddot{\boldsymbol{s}} = -\omega^2 \boldsymbol{s}$ und $\dot{\boldsymbol{s}} = \boldsymbol{\omega} \times \boldsymbol{s}$ einsetzt.

Die Grenzfälle sind bekannt. Für $\omega_0 = 0$ liegt eine Bewegung mit der Zyklotronfrequenz $\omega = 2\omega_L$ im homogenen Magnetfeld vor (vgl. 2.1.7.1.). Für $\boldsymbol{B} = 0$ verbleibt die harmonische Schwingung (vgl. 2.1.5.1.) $\omega = \pm\omega_0$.

Im schwachen Magnetfeld $\omega_L \ll \omega_0$ erhält man $\omega = \pm\omega_0 + \omega_L$. Der Bewegung im Zentralkraftfeld wird die Larmor-Präzession, eine Drehung mit der Frequenz ω_L, überlagert. In dieser Form gilt diese Aussage für beliebige Zentralkraftfelder. Aus der Larmor-Präzession läßt sich der Zeeman-Effekt, d. h. die Aufspaltung der Spektrallinien im Magnetfeld, verstehen.

2.1.8. Verallgemeinerte Formen der Bewegungsgleichung

Mit der Newtonschen Grundgleichung liegt die Bewegungsgleichung der Mechanik vor. Durch Integration kann man die Bahnkurve gewinnen. Eine Verallgemeinerung trägt folgendem Anliegen Rechnung: Es werden nur so viel Koordinaten eingeführt, wie das System *Freiheitsgrade* aufweist. Zwangsbedingungen, wie Aufhängungen, starre Verbindungen zwischen Massen, schränken die Bewegungsmöglichkeiten ein und üben *Zwangskräfte* auf die Massen aus. So zieht der Pendelfaden an der Pendelmasse. Diese Zwangskräfte sind durch die Zwangsbedingungen und durch den Bewegungsablauf bestimmt. Man kann sie nur parallel zur Lösung der Bewegungsgleichung gewinnen. Es gelingt aber, sie in verallgemeinerten Formulierungen zu eliminieren. Weiterhin sind in den verallgemeinerten Formulierungen Koordinatentransformationen relativ leicht durchführbar.

Darüber hinaus soll die theoretische Mechanik als *Modelltheorie* für andere Gebiete der Physik fungieren. Es gibt Gesetzmäßigkeiten in der Natur, die in vielen oder allen Gebieten der Physik gelten. Sie werden also auch in der Mechanik gefunden. Es ist daher von der Mechanik ausgehend zu fragen, welche Ergebnisse über die Mechanik hinaus gelten. Die Verallgemeinerung im Lagrange- und Hamilton-Formalismus führt auf eine Gestalt der Bewegungsgleichungen, in die sich auch Bewegungsgleichungen anderer Gebiete der Physik bringen lassen. Die Schlußfolgerungen, die man mittels des Hamilton-Formalismus in der Mechanik gewinnt, sind also auch für diese anderen Gebiete zutreffend.

2.1.8.1. Lagrange-Formalismus. Das Variationsprinzip $\delta \int_{t_1}^{t_2} \mathrm{d}t L(q_1, \ldots, q_f; \dot{q}_1, \ldots, \dot{q}_f; t) = 0$ ist *der Eulerschen Differentialgleichung* (vgl. 13.3.3.10.) $\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{L}{\partial q_k} = 0$, $k = 1, 2, \ldots, f$ äquivalent. Findet man nun ein solches Funktional L, daß die Eulersche Gleichung mit der Newtonschen Bewegungsgleichung identisch ist, dann hat man die Grundgleichung der Mechanik als Variationsprinzip formuliert. Man nennt L dann die *Lagrange-Funktion* des betrachteten Systems und die Eulersche Differentialgleichung Lagrange-Gleichung (II. Art).

Der Wert des Integrals über L ändert sich bei einer Koordinatentransformation $q_k(t) = q_k[q_l(t), t]$ nicht. Die Aussage des Variationsprinzips gilt unabhängig von der Koordinatenwahl. Damit haben die Bewegungsgleichungen auch unabhängig von der Koordinatenwahl die gleiche Gestalt, es sind die Lagrange-Gleichungen.

Für ein System, in dem die Kräfte durch Potentiale beschreibbar sind, ist die Lagrange-Funktion gleich der Differenz $L = T - V$, kinetische minus potentielle Energie.

Das Aufstellen der *Bewegungsgleichungen* erfolgt also auf folgendem Wege: 1. Man bestimmt die Zahl der Freiheitsgrade f des betrachteten Systems. 2. Man wählt f Koordinaten, die den Bewegungsablauf am zweckmäßigsten beschreiben. 3. Man bestimmt die Lagrange-Funktion, indem man kinetische und potentielle Energie in diesen Koordinaten ausdrückt. Alle Eigenschaften des Systems werden durch eine Funktion L charakterisiert. 4. Die Lagrange-Gleichungen II. Art sind die Bewegungsgleichungen.

Ein *ebenes Pendel* (vgl. Abb. 2.1.5–2) hat einen Freiheitsgrad. Die zweckmäßigste Koordinate ist ϑ. Die Geschwindigkeit der Masse ist $l\dot{\vartheta}$. Bei einer Auslenkung ϑ wird die Masse um $l(1 - \cos \vartheta)$ angehoben. Daher ist die Lagrange-Funktion $L = (m/2)(l\dot{\vartheta})^2 - mgl(1 - \cos \vartheta)$. Aus ihr folgt die Bewegungsgleichung $ml^2\ddot{\vartheta} + mgl \sin \vartheta = 0$, bei kleinen Auslenkungen $\ddot{\vartheta} + (g/l)\vartheta = 0$.

Das *sphärische Pendel* (Abb. 2.1.8–1) hat zwei Freiheitsgrade. Die Lagrange-Funktion enthält die Koordinate φ nicht, $L = (m/2)l^2(\dot{\vartheta}^2 + \sin^2 \vartheta\dot{\varphi}^2) - mgl(1 - \cos \vartheta)$. Daher ersieht man unmittelbar aus der Lagrange-Gleichung für φ, daß die Drehimpulskomponente $L_z = ml^2 \sin^2 \vartheta\dot{\varphi}$ eine Erhaltungsgröße ist. Die Bewegungsgleichung für ϑ lautet bei kleinen Auslenkungen $\ddot{\vartheta} + (g/l)\vartheta = \vartheta\dot{\varphi}^2$. Neben der Schwingung $\dot{\varphi} = 0$ wie beim ebenen Pendel ist $\dot{\varphi} =$ konst.,

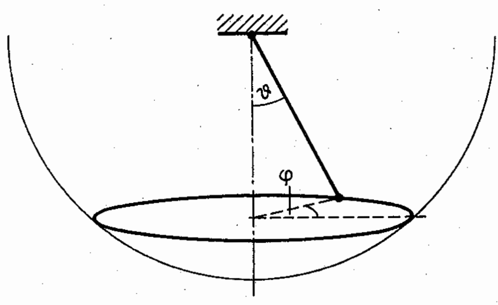


Abb. 2.1.8–1 Sphärisches Pendel

ϑ = konst., also eine Bewegung mit konstanter Geschwindigkeit auf einem Kreis um die Vertikale, eine spezielle Bewegungsform des sphärischen Pendels.
Bewegt sich eine Masse auf einer *schiefen Ebene*, die mit der Geschwindigkeit $v(t)$ (Abb. 2.1.8-2) verschoben wird, dann ist $L = (m/2)\,[(\dot{s}\cos\alpha + v)^2 + (\dot{s}\sin\alpha)^2] + mgs\sin\alpha$. Dazu gehört die Bewegungsgleichung $\ddot{s} = g\sin\alpha - \dot{v}\cos\alpha$. Bei $\dot{v} > g\tan\alpha$ wird sich eine anfangs ruhende Masse die Ebene hinauf bewegen.
Eine Lagrange-Funktion läßt sich auch finden, wenn die Kräfte kein Potential haben. Eine geschwindigkeitsproportionale Reibung erhält man mit $L = L_0 e^{\gamma t}$, wenn L_0 die Lagrange-Funktion für den ungedämpften Fall ist.

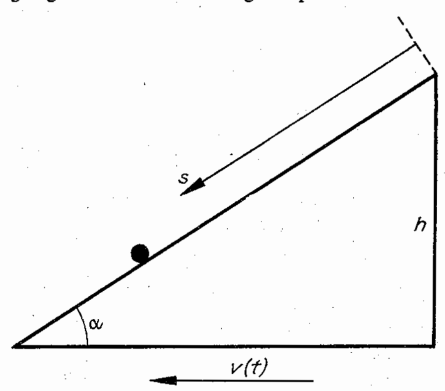


Abb. 2.1.8–2 Bewegte schiefe Ebene

Für die Bewegung im *elektromagnetischen Feld* $\boldsymbol{E}(\boldsymbol{r}, t) = -\nabla U(\boldsymbol{r}, t) - \partial \boldsymbol{A}(\boldsymbol{r}, t)/\partial t$, $\boldsymbol{B}(\boldsymbol{r}, t) = \nabla \times \boldsymbol{A}(\boldsymbol{r}, t)$ führt $L = (m/2)\dot{\boldsymbol{s}}^2 + q\dot{\boldsymbol{s}}\cdot\boldsymbol{A} - qU$ auf die Bewegungsgleichung $m\ddot{\boldsymbol{s}} = q(\boldsymbol{E} + \dot{\boldsymbol{s}} \times \boldsymbol{B})$.

2.1.8.2. Hamilton-Formalismus. Beim Übergang zum Hamilton-Formalismus werden die Vorteile des Lagrange-Formalismus (Charakterisierung des Systems durch eine Funktion, beliebige Wählbarkeit der Koordinaten, einheitliche Form der Bewegungsgleichungen, leichte Erkennbarkeit von Erhaltungsgrößen) beibehalten. Darüber hinaus werden die f Gleichungen zweiter Ordnung in $2f$ Differentialgleichungen erster Ordnung umgeformt, die schon nach den Zeitableitungen aufgelöst sind. Die Hamilton-Funktion hat eine anschauliche Bedeutung, sie ist gleich der Energie (falls diese eine Erhaltungsgröße ist). Die Form der Bewegungsgleichungen ist gegen eine größere Gruppe von Transformationen invariant.
Die Hamilton-Funktion wird aus der Lagrange-Funktion berechnet, $H(p_k, q_k, t) = \sum_k \dot{q}_k \dfrac{\partial L}{\partial \dot{q}_k} - L$, wobei die Geschwindigkeiten $\dot{q}_k$ zu eliminieren sind, indem man die über $p_k = \partial L/\partial \dot{q}_k$ definierten *kanonisch konjugierten Impulse* einführt. Die Bewegungsgleichungen sind dann die *kanonischen Gleichungen* $\dot{p}_k = -\dfrac{\partial H}{\partial q_k}$, $\dot{q}_k = \dfrac{\partial H}{\partial p_k}$, $H = H(q_k, p_k, t)$.
Haben die Kräfte in einem System ein Potential, dann ist die Hamilton-Funktion die Summe von kinetischer und potentieller Energie $H = T + V = W$, also gleich der mechanischen Energie W. In diesem Fall ist bei der Wahl kartesischer Koordinaten der kanonische Impuls gleich dem mechanischen Impuls, für ein Teilchen ist also $H = \boldsymbol{p}^2/2m + V(\boldsymbol{r})$. Wählt man als Koordinate einen Winkel, der die Drehung um eine Achse beschreibt, dann ist der kanonisch konjugierte Impuls gleich der Drehimpulskomponente um diese Achse. Für die Bewegung im Magnetfeld führt die Lagrange-Funktion (vgl. 2.1.8.1.) auf den kanonisch konjugierten Impuls $\boldsymbol{p} = m\dot{\boldsymbol{r}} + q\boldsymbol{A}$. Die Hamilton-Funktion lautet also $H = (\boldsymbol{p} - q\boldsymbol{A})^2/2m + qU$.
Für die in einer Ebene erfolgende Bewegung in einem Zentralkraftfeld folgen dann aus der Lagrange-Funktion $L = (m/2)\,(\dot{r}^2 + r^2\dot{\varphi}^2) - V(r)$ die kanonischen Impulse $p_r = m\dot{r}$ und $p_\varphi = mr^2\dot{\varphi} = |\boldsymbol{L}|$. Daraus ergibt sich die Hamilton-Funktion $H = p_r^2/2m + L^2/2mr^2 + V(r)$. Da H von φ unabhängig ist, liefert die kanonische Gleichung $\dot{L} = 0$, der Betrag des Drehimpulses ist eine Erhaltungsgröße. Das Effektivpotential (vgl. 2.1.6.2.) taucht sofort in der Bewegungsgleichung für $r(t)$ auf.

2.1.8.3. Zwangskräfte. Aufhängungen, Schienen, starre Verbindungen usw. üben Zwangskräfte auf die Massen aus. In den Lagrange-Gleichungen II. Art tauchen diese Zwangskräfte nicht mehr auf. Gelegentlich ist ihre Größe aber gefragt.
Virtuelle Verrückungen $\delta\boldsymbol{r}$ sind solche Verschiebungen der Massen, die trotz der Zwangsbedingungen zu einem festgehaltenen Zeitpunkt möglich sind. Sind die Zwangsbedingungen zeitunabhängig (skleronom), dann sind virtuelle Verrückungen tatsächlich mögliche Verschiebungen. Das d'Alambertsche Prinzip besagt (als Erfahrungsaussage bei Vernachlässigung von Reibungskräften), daß die Zwangskräfte senkrecht zu den virtuellen Verrückungen stehen, $\boldsymbol{Z}\cdot\delta\boldsymbol{r} = 0$. Bei zeitunabhängigen Zwangsbedingungen leisten die Zwangskräfte keine Arbeit. Kann man die Zwangsbedingungen z. B. dadurch charakterisieren, daß sich das Teilchen auf einer Fläche $f(\boldsymbol{r}, t) = 0$ bewegen soll (beim Pendel ist z. B. $r - l = 0$), dann ist die Richtung der Kraft durch die Flächennormale $\nabla f(\boldsymbol{r}, t)$ gegeben. Die Zwangskraft hat die Form $\boldsymbol{Z} = \lambda\nabla f$. Die Newtonsche Bewegungsgleichung läßt sich in der Form der *Lagrange-Gleichung I. Art* schreiben $m\ddot{\boldsymbol{s}} = \boldsymbol{F} + \lambda\nabla f$, $f(\boldsymbol{r}, t) = 0$.
Bei der schiefen Ebene (vgl. 2.1.4.3.) ist z. B. $\boldsymbol{F} = m\boldsymbol{g}$ die Schwerkraft und $\boldsymbol{Z}$ das Negative der Komponente der Schwerkraft senkrecht zur Ebene. Bewegungsgleichung und Zwangsbedingung bilden ein Gleichungssystem, aus dem man $\boldsymbol{s}(t)$ und λ bestimmen kann.

2.1.8.4. Invarianzeigenschaften und Erhaltungsgrößen. Die Erfahrungen zeigen, daß in unserer Umwelt kein Zeitpunkt, kein Raumpunkt, keine Raumrichtung und kein Inertialsystem ausgezeichnet ist. So kann man z. B. voraussetzen, daß ein Experiment unter gleichen Bedingungen im Jahre 2000 zu den gleichen Ergebnissen führt wie im Jahre 1980 und daß die Resultate unabhängig davon sind, in welcher Stadt der Versuch wiederholt wird.
Kann man ein System durch eine Hamilton-Funktion charakterisieren, d. h. existiert eine Funktion, die alle Eigenschaften des Systems enthält und aus der man die Bewegungsgleichungen als kanonische Gleichungen gewinnen kann, dann muß diese Funktion die oben genannten Symmetrieeigenschaften widerspiegeln.
In einem System, das äußeren Einflüssen unterliegt, können Zeitpunkte, Raumpunkte oder Raumrichtungen ausgezeichnet sein. Bei der Wurfbewegung z. B. ist durch die Schwerkraft eine Raumrichtung ausgezeichnet. In einem abgeschlossenen System ist dies aber nicht der Fall. Dann können folgende Aussagen abgeleitet werden: Aus der Invarianz der Hamilton-Funktion gegen *Translationen der Zeit* ($t \to t + dt$) folgt, daß die Energie eine Erhaltungsgröße ist.

Dabei repräsentiert H die Energie. Aus der Invarianz der Hamilton-Funktion gegen *Translationen des Raumes* ($\boldsymbol{r} \rightarrow \boldsymbol{r} + \mathrm{d}\boldsymbol{r}$) folgt, daß der Impuls eine Erhaltungsgröße ist. Aus der Invarianz der Hamilton-Funktion gegen *Drehungen des Raumes* ($\boldsymbol{r} \rightarrow \boldsymbol{r} + \mathrm{d}\vec{\varphi} \times \boldsymbol{r}$) folgt, daß der Drehimpuls eine Erhaltungsgröße ist. Aus der Invarianz gegen den *Wechsel des Inertialsystems* ($\boldsymbol{v} \rightarrow \boldsymbol{v} + \mathrm{d}\boldsymbol{v}$) folgt, daß der Anfangsschwerpunkt eine Erhaltungsgröße ist.

In einem abgeschlossenen System existieren also mindestens diese zehn Erhaltungsgrößen. Dies folgt allein aus der Existenz einer Hamilton-Funktion. Diese Invarianzforderungen beinhalten auch Aussagen über die *Kräfte zwischen den Teilchen* eines mechanischen Systems. So kann der Impuls nur dann eine Erhaltungsgröße sein, wenn $\boldsymbol{F}_{21} = -\boldsymbol{F}_{12}$ – also das dritte Newtonsche Axiom – gilt. Ist die Paarwechselwirkung durch ein nur ortsabhängiges Potential beschreibbar, so darf dieses nur vom Betrag des Abstandes beider Teilchen abhängen, $V = V(|\boldsymbol{r}_1 - \boldsymbol{r}_2|)$. Auch die Form von spin- und impulsabhängigen Kräften, wie sie bei der Nukleon-Nukleon-Wechselwirkung auftreten, ist weitgehend festgelegt.

Auch ein *nichtabgeschlossenes System* kann noch Invarianzeigenschaften aufweisen. Sind die äußeren Kraftfelder zeitunabhängig, dann tritt in der Hamilton-Funktion die Zeit nicht explizit auf, die Energie ist eine Erhaltungsgröße. In einem Zentralkraftfeld ist der Drehimpuls eine Erhaltungsgröße.

2.1.8.5. Mehrteilchensysteme. Man betrachtet ein System von Massen m_n, zwischen denen paarweise Wechselwirkungen mit einem Potential $V_{nm}(|\boldsymbol{r}_n - \boldsymbol{r}_m|)$ bestehen. Weiterhin wirkt auf jedes Teilchen eine äußere Kraft, das zugehörige Kraftfeld ist jeweils durch das Potential $V_{a,n}(\boldsymbol{r})$ beschreibbar. Dann ist die Hamilton-Funktion

$$H = \sum_n \left(\frac{p_n^2}{2m_n} + V_{a,n}(\boldsymbol{r}_n) \right) + \sum_{n<m} V_{nm}(|\boldsymbol{r}_n - \boldsymbol{r}_m|).$$

Liegen gleichartige Teilchen vor, dann hängen die Potentiale nicht mehr von der Teilchennummer ab,

$$H = \sum_n \left(\frac{p_n^2}{2m_n} + V_a(\boldsymbol{r}_n) \right) + \sum_{n<m} V(|\boldsymbol{r}_n - \boldsymbol{r}_m|).$$

So wäre die Hamilton-Funktion für die Elektronen in einem Atom (die Besonderheiten der quantenmechanischen Betrachtung vgl. 4.)

$$H = \sum_{n=1}^{Z} \left(\frac{p_n^2}{2m} - \frac{Z\epsilon^2}{r_n} \right) + \sum_{n<m}^{1\ldots Z} \frac{\epsilon^2}{|\boldsymbol{r}_n - \boldsymbol{r}_m|},$$

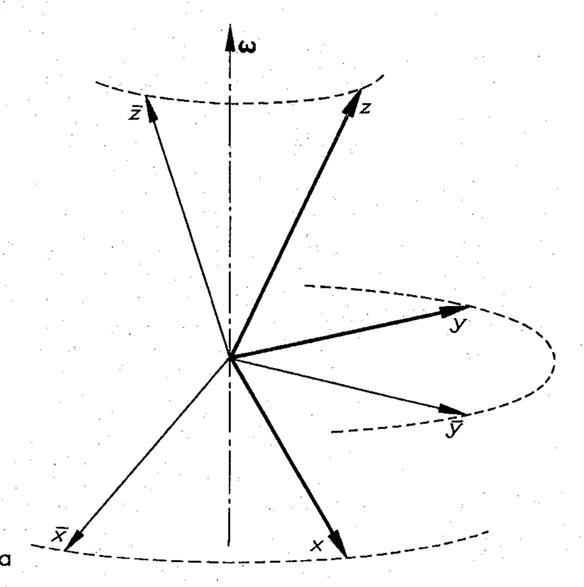
wobei der Koordinatenursprung in den Atomkern gelegt und der Atomkern als ruhend angenommen wurde. Flüssigkeiten oder Festkörper, die nur eine chemische Komponente besitzen, bestehen aus zwei Teilchenarten, den Z-fach geladenen Kernen (N Kerne) und den NZ Elektronen. In der Hamilton-Funktion

$$H = \sum_{l=1}^{N} \frac{p_l^2}{2M} + \sum_{n=1}^{NZ} \left(\frac{p_n^2}{2m} - \sum_{l=1}^{N} \frac{Z\epsilon^2}{|\boldsymbol{r}_n - \boldsymbol{R}_l|} \right) + \sum_{n<m}^{1\ldots NZ} \frac{\epsilon^2}{|\boldsymbol{r}_n - \boldsymbol{r}_m|} + \sum_{l<k}^{1\ldots N} \frac{Z^2\epsilon^2}{|\boldsymbol{R}_l - \boldsymbol{R}_k|}$$

bedeuten $\boldsymbol{R}_l$ die Orte der Kerne und $\boldsymbol{r}_n$ die Orte der Elektronen.

2.1.9. Rotierendes Bezugssystem

Absolutgeschwindigkeiten sind nicht meßbar, die Newtonsche Grundgleichung ist invariant gegen Galilei-Transformationen, d. h. gegen den Übergang zu einem Bezugssystem, das sich zum Ausgangssystem mit konstanter Geschwindigkeit bewegt, sie hat in allen Inertialsystemen die gleiche Form. Führt man jedoch Beobachtungen in einem Bezugssystem aus, das sich zu einem Inertialsystem beschleunigt bewegt, dann gewinnt man eine andere Bewegungsgleichung. Der Unterschied in den Beschleunigungen wird als Scheinkraft, als *Trägheitskraft* registriert. Ein Laborsystem auf der Erde führt die Drehung um die Erdachse



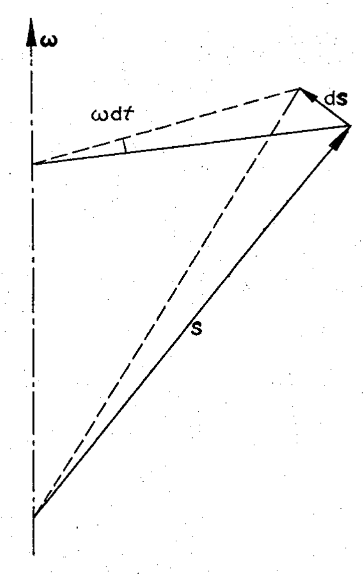


Abb. 2.1.9–1 (a) rotierendes Bezugssystem und (b) Richtung und Betrag von d$\boldsymbol{s}$ bei einer Drehung

mit aus, ist also ein rotierendes Bezugssystem und damit ein beschleunigtes Bezugssystem. Eine Bewegung auf einem Kreis ist ja eine beschleunigte Bewegung. Wird eine Masse an einem Faden auf einer Kreisbahn geführt, dann muß man über den Faden eine *Zentripedalkraft* auf die Masse ausüben, die gleich Masse mal Beschleunigung der Masse ist. In einem mit derselben Winkelgeschwindigkeit rotierenden System ruht die Masse. Will man mit der Newtonschen Gleichung in gleicher Form wie im Inertialsystem arbeiten, muß man sich auf den Standpunkt stellen, daß die Masse sich in einem Kraftfeld bewegt, das durch die über den Faden übertragene Kraft kompensiert wird. Dieses scheinbare Kraftfeld, in diesem Fall die Zentrifugalkraft, entspricht der Differenz der Beschleunigungen in beiden Bezugssystemen. Da viele Messungen in Laborsystemen ausgeführt werden, die sich durch die Erddrehung auf einem Kreis bewegen, diese Drehung aber oft nicht bewußt berücksichtigt wird, muß man sich fragen, welchen Einfluß die Trägheitskräfte auf die Beobachtungen haben.

2.1.9.1. Trägheitskräfte. Ein Punkt der Bahnkurve $\bar{s}(t)$ in einem (raumfesten) Inertialsystem ist in kartesischen Koordinaten durch $\bar{s}(t) = \bar{x}(t)\bar{e}_x + \bar{y}(t)\bar{e}_y + \bar{z}(t)\bar{e}_z$ gegeben. In einem dazu mit der Winkelgeschwindigkeit ω (Abb. 2.1.9–1) rotierenden System lautet die Komponentenzerlegung des gleichen Vektors, der in diesem System aber $s(t)$ genannt werden soll, $s(t) = x(t)e_x(t) + y(t)e_y(t) + z(t)e_z(t)$. Der wesentliche Unterschied zwischen beiden Darstellungen liegt darin, daß die Basisvektoren im rotierenden System zeitabhängig sind. Unter $\dot{s}(t)$ soll jetzt aber nicht die Zeitableitung von $s(t)$ verstanden werden, sondern es wird definiert $\dot{s}(t) = \dot{x}(t)e_x + \dot{y}(t)e_y + \dot{z}(t)e_z$. Das ist die Geschwindigkeit, die ein Beobachter im rotierenden System feststellt. Ebenso wird definiert $\ddot{s}(t) = \ddot{x}(t)e_x + \ddot{y}(t)e_y + \ddot{z}(t)e_z$. Bei der vollständigen Zeitableitung muß man die Zeitabhängigkeit der Einheitsvektoren berücksichtigen. Wie Abb. 2.1.9–1 zeigt, ändern sie sich bei einer Drehung um den Winkel $\omega\,dt$ um $de = \omega\,dt \times e$, d. h., es ist $\dot{e} = \omega \times e$. In dieser Weise transformieren sich alle Vektoren bei einer Drehung. Wie vermerkt, ist $\bar{s}(t) = s(t)$. Für die Geschwindigkeiten gilt dann die Gleichung

$$\dot{\bar{s}}(t) = \dot{s} + \omega \times s.$$

Ein im rotierenden Bezugssystem ruhender Punkt ($\dot{s} = 0$) bewegt sich im Inertialsystem mit der Geschwindigkeit $\omega \times s$, d. h. auf einem Kreis.

In der Lagrange-Funktion muß die kinetische Energie

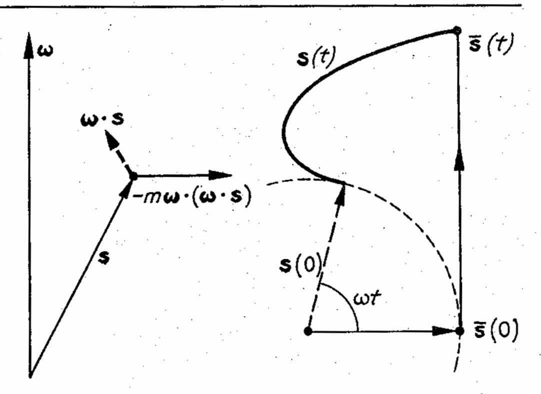


Abb. 2.1.9–2 Zentrifugalkraft (*links*)

Abb. 2.1.9–3 Geradlinige Bewegung $\bar{s}(t)$ (———) vom rotierenden System aus gesehen (– – –). ω zeigt aus der Papierebene heraus

durch $T = (m/2)\dot{\bar{s}}^2 = (m/2)\,(\dot{s} + \omega \times s)^2$ ausgedrückt werden. Daraus folgt

$$m\ddot{s} = F - m\omega \times (\omega \times s) - 2m\omega \times \dot{s} - m\dot{\omega} \times s$$

als Bewegungsgleichung. Die neben F stehenden Ausdrücke sind die Trägheitskräfte, die Scheinkräfte. Der erste Term ist die *Zentrifugalkraft*. Sie zeigt radial nach außen (Abb. 2.1.9–2) und ist betragsmäßig $m\omega^2 a$, mit a als Abstand von der Drehachse. Den zweiten Term nennt man *Coriolis-Kraft*. Er tritt nur für $\dot{s} \neq 0$ auf, d. h., wenn sich das Teilchen im rotierenden Bezugssystem bewegt. Ihre Bedeutung kann man veranschaulichen, wenn man eine Masse betrachtet, die sich im Inertialsystem mit konstanter Geschwindigkeit auf einer Geraden bewegt (Abb. 2.1.9-3). Im bewegten Bezugssystem erhält man eine relativ komplizierte Bahn. Die Coriolis-Kraft bewirkt eine Rechtsabweichung. Es tritt also bei einer Bewegung auf der Erdoberfläche eine Abweichung nach Osten auf, wenn sich die Masse auf der nördlichen Halbkugel anfangs in Richtung Nordpol oder auf der südlichen Halbkugel in Richtung Südpol bewegt.

Der dritte Term ist nur dann vorhanden, wenn sich die Winkelgeschwindigkeit im Zeitablauf ändert. Er soll nicht diskutiert werden.

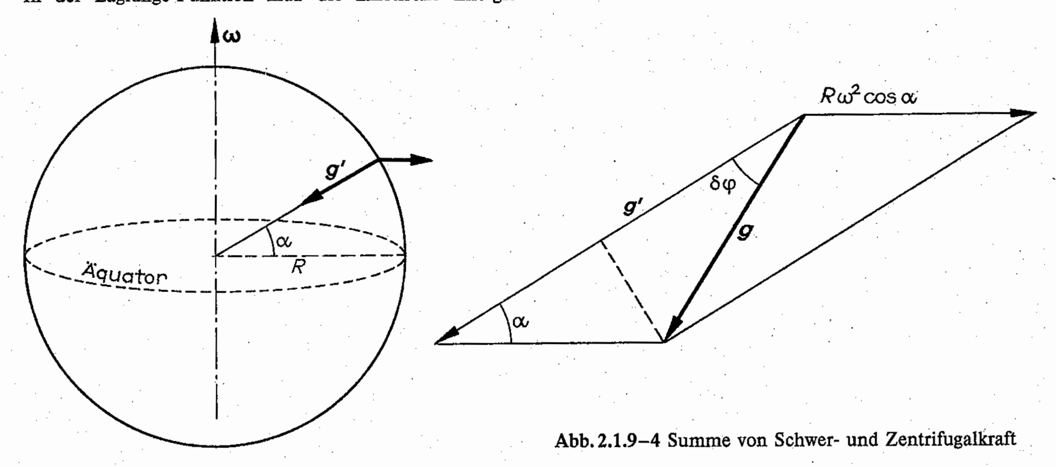


Abb. 2.1.9–4 Summe von Schwer- und Zentrifugalkraft

2.1.9.2. Einfluß der Erdrotation. Die Erde rotiert mit der Kreisfrequenz $\omega = 2\pi/86\,162\,\mathrm{s}^{-1} = 7{,}29 \cdot 10^{-5}\mathrm{s}^{-1}$, die sich aus der Dauer eines Sterntages ergibt. Das ist eine kleine Frequenz, daher spürt man von den Trägheitskräften nur wenig. Auf eine auf der Erdoberfläche ruhende Masse wirkt die Schwerkraft $m\boldsymbol{g}'$ und als Trägheitskraft die Zentrifugalkraft. Beide setzen sich zu einem Vektor $\boldsymbol{g}$ zusammen, $\boldsymbol{g} = \boldsymbol{g}' - \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{s})$, der nicht zum Erdmittelpunkt hinweist (Abb. 2.1.9–4). Die Stärke der Abplattung der Erde ($\approx 1/297$) läßt sich aus der Bedingung berechnen, daß sich eine Oberfläche einstellt, auf der $\boldsymbol{g}$ überall senkrecht steht. Eine Tangentialkomponente der Kraft würde Massenverschiebungen bewirken, der stationäre Gleichgewichtszustand hätte sich noch nicht eingestellt.

Der Vektor $\boldsymbol{g}$, der als Schwerebeschleunigung auf der Erde bezeichnet wird, schließt die Zentrifugalkraft mit ein. $\boldsymbol{g}'$ ist aus der Massenverteilung der abgeplatteten Erde zu berechnen. Bestimmt man $\boldsymbol{g}'$ näherungsweise aus einer kugelsymmetrischen Massenverteilung, dann kann man aus der Abb. 2.1.9–4 $g - g' = -\omega^2 R \cos^2\alpha$ und $\delta\varphi = (\omega^2 R/2g)\sin 2\alpha$ ablesen. Bei der genauen Betrachtung ändern sich die Vorfaktoren dieser Korrekturen (vgl. 2.1.3.3.). Es ist aber zu sehen, daß durch die Erdrotation Zusatzterme auftreten, deren Größenordnung durch den kleinen dimensionslosen Parameter $\omega^2 R/g = 0{,}003\,5$ bestimmt wird.

2.1.9.3. Fall und Wurf. Die Bewegungsgleichung $\ddot{\boldsymbol{s}} = \boldsymbol{g} - 2\boldsymbol{\omega} \times \dot{\boldsymbol{s}}$ enthält nur die Coriolis-Kraft, da die Zentrifugalkraft in $\boldsymbol{g}$ erfaßt ist und $\boldsymbol{\omega}$ zeitlich konstant ist. Bei Beschränkung auf Korrekturen in erster Ordnung in ω kann man in die Coriolis-Kraft die sich für $\omega = 0$ ergebende Bahnkurve einsetzen. Man erhält $\boldsymbol{s}(t) = \boldsymbol{g}t^2/2 + \boldsymbol{v}t + \boldsymbol{s}_0 + \boldsymbol{u}(t)$, $\boldsymbol{u}(t) = -\boldsymbol{\omega} \times \boldsymbol{v}t^2 - \boldsymbol{\omega} \times \boldsymbol{g}t^3/3$. Beim Fall aus der Höhe h ($\boldsymbol{v} = 0$, $\boldsymbol{s}_0 = h\boldsymbol{e}_z$, Abb. 2.1.9–5) ergibt sich auf der Nordhalbkugel eine Ostabweichung. Die Masse trifft nicht bei $x = 0$ sondern bei $x = (2^{3/2}/3)\sqrt{(\omega^2 R/g)\,h/R}\;h\cos\alpha$ auf. Bei einer Fallhöhe von 100 m liegt die Abweichung in der Größenordnung 1 cm.

2.1.9.4. Foucaultsches Pendel. Für das sphärische Pendel beobachtet man in einem sich mit der Erde mitdrehenden Bezugssystem nicht die in 2.1.8.1. beschriebene Lösung. Die Coriolis-Kraft führt zu Zusatztermen in den Bewegungsgleichungen $\frac{\mathrm{d}}{\mathrm{d}t}[\vartheta^2(\dot\varphi + \omega\sin\alpha)] = 0$, $\ddot\vartheta + (g/l)\vartheta = \vartheta(\dot\varphi + \omega\sin\alpha)$, die für $\omega = 0$ verschwinden. Eine einfache Lösung ist erkennbar. Zur Schwingung des mathematischen Pendels $\ddot\vartheta + (g/l)\vartheta = 0$ gehört $\dot\varphi = -\omega\sin\alpha$. Die Pendelebene dreht sich um $15\sin\alpha$ Grad je Stunde. Foucault konnte bei seinen um 1850 ausgeführten Versuchen diese Abhängigkeit vom Breitengrad nachweisen.

2.1.9.5. Bewegung im Magnetfeld. Beeinflußt man die Bewegung in einem Kraftfeld $\boldsymbol{F}$ durch ein schwaches Magnetfeld, dann lautet die Bewegungsgleichung $m\ddot{\boldsymbol{s}} = \boldsymbol{F} + q\dot{\boldsymbol{s}} \times \boldsymbol{B}$, in einem mit der Kreisfrequenz $\boldsymbol{\omega}$ rotierenden System $m\ddot{\boldsymbol{s}} = \boldsymbol{F} + q\dot{\boldsymbol{s}} \times \boldsymbol{B} + q(\boldsymbol{\omega} \times \boldsymbol{s}) \times \boldsymbol{B} - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \boldsymbol{s}) - 2m\boldsymbol{\omega} \times \dot{\boldsymbol{s}}$. Wählt man $\boldsymbol{\omega}$ gleich der Larmor-Frequenz $\boldsymbol{\omega}_L = -q\boldsymbol{B}/2m$ und vernachlässigt in $\boldsymbol{B}$ quadratische Terme, dann ergibt sich im rotierenden System die gleiche Bewegungsgleichung $m\ddot{\boldsymbol{s}} = \boldsymbol{F}$ wie ohne Magnetfeld im ruhenden. Der Einfluß des Magnetfeldes besteht also in erster Ordnung darin, daß die Bahnkurven mit der Larmor-Frequenz gedreht werden. Bahndeformationen treten erst in zweiter Ordnung auf.

2.1.10. Zweikörperproblem

Betrachtet wird die Bewegung zweier Massen unter dem Einfluß ihrer gegenseitigen Wechselwirkung. Äußere Kräfte sollen nicht einwirken, so daß ein abgeschlossenes System vorliegt. Diese Situation ist z. B. vorhanden, wenn man bei der Planetenbewegung die Mitbewegung der Sonne berücksichtigen will. Das Zweikörperproblem ist der einfachste Fall eines Mehrkörperproblems. In ihm können analytische Lösungen angegeben werden. Beim Dreikörperproblem – wie auch bei Systemen mit mehr Teilchen – sind nur Näherungslösungen oder numerische Lösungen möglich.

2.1.10.1. Erhaltungsgrößen. Für ein *abgeschlossenes System* sind mit Energie, Impuls, Drehimpuls und Anfangsschwerpunkt zehn Erhaltungsgrößen bekannt. Da bei der Lösung der Newtonschen Bewegungsgleichungen für zwei Teilchen

Abb. 2.1.9–5 Ostabweichung beim Fall infolge Erddrehung

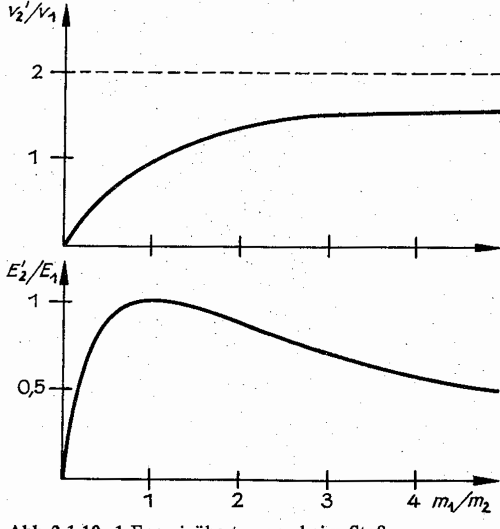


Abb. 2.1.10–1 Energieübertragung beim Stoß

insgesamt zwölf Integrationskonstanten auftreten, z. B. zwei Anfangsorte und zwei Anfangsgeschwindigkeiten, ist der Bewegungsablauf durch Angabe der zehn Erhaltungsgrößen schon weitestgehend charakterisiert.

Durch die Erhaltungsgrößen und die Aussage, daß das Potential der gegenseitigen Wechselwirkung nur vom Betrag des Abstandes der beiden Massen abhängen kann, werden *Schwerpunkt-* und *Differenzvektor* als zweckmäßige Koordinaten nahegelegt, $\boldsymbol{s} = (m_1\boldsymbol{r}_1 + m_2\boldsymbol{r}_2)/(m_1 + m_2)$, $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$. Nach Einführung der Begriffe *Gesamtmasse* $M = m_1 + m_2$ und *reduzierte Masse* $\mu = m_1 m_2/(m_1 + m_2)$ lauten die zu diesen Koordinaten gehörenden kanonischen Impulse $\boldsymbol{P} = \boldsymbol{p}_1 + \boldsymbol{p}_2 = M\dot{\boldsymbol{s}}$ und $\boldsymbol{p} = \mu\dot{\boldsymbol{r}}$. $\boldsymbol{p} = m_1(\boldsymbol{p}_1/m_1 - \boldsymbol{P}/M)$ ist der Impuls der Masse m_1 im *Schwerpunktsystem*, in dem also die Geschwindigkeit um die Schwerpunktgeschwindigkeit verringert ist.

Die Masse m_2 hat im Schwerpunktsystem den Impuls $-\boldsymbol{p}$, definitionsgemäß ist der Gesamtimpuls im Schwerpunktsystem Null. Aus der Hamilton-Funktion

$$H = \frac{\boldsymbol{p}_1^2}{2m_1} + \frac{\boldsymbol{p}_2^2}{2m_2} + V(|\boldsymbol{r}_1 - \boldsymbol{r}_2|) = \frac{\boldsymbol{P}^2}{2M} + \frac{\boldsymbol{p}^2}{2\mu} + V(r)$$

erkennt man, daß Schwerpunkt- und Relativbewegung entkoppeln. Der Schwerpunkt bewegt sich auf einer Geraden $\boldsymbol{s}(t) = (\boldsymbol{P}/M)t + \boldsymbol{s}_0$. Die Relativbewegung entspricht der Bewegung eines Teilchens mit der Masse μ im Zentralpotential $V(r)$.

Die in 2.1.6. zusammengestellten Aussagen können übernommen werden. Im Falle der Planetenbewegung ergeben sich Ellipsenbahnen um den gemeinsamen Schwerpunkt in einer Ebene senkrecht zum Drehimpuls. Die Ausmaße der Ellipsen der beiden Massen verhalten sich wie m_1/m_2.

2.1.10.2. Zentraler Stoß. Beim zentralen Stoß bewegen sich zwei Massen mit Geschwindigkeiten v_1 und v_2 auf einer Geraden. Die Geschwindigkeiten nach dem Stoß v_1' und v_2' sind zu bestimmen. Da keine äußeren Kräfte wirken sollen, ist der *Impuls eine Erhaltungsgröße*, der Impuls nach dem Stoß ist gleich dem Impuls vor dem Stoß $m_1v_1' + m_2v_2' = m_1v_1 + m_2v_2$. *Die mechanische Energie ist eine Erhaltungsgröße, wenn der Stoß elastisch erfolgt.* Sonst wird ein Teil der mechanischen Energie für eine bleibende Verformung oder für eine Erwärmung der stoßenden Körper aufgebraucht. Mit $\alpha \leqq 1$ schreibt man $m_1v_1'^2/2 + m_2v_2'^2/2 = \alpha(m_1v_1^2/2 + m_2v_2^2/2)$. $\alpha = 1$ entspricht dem elastischen Stoß.

Für den *elastischen Stoß* folgt aus Impuls und Energiebilanz

$$v_1' = v_1 + \frac{2m_2}{m_1 + m_2}(v_2 - v_1), \; v_2' = v_2 + \frac{2m_1}{m_1 + m_2}(v_1 - v_2).$$

Die Impulsübertragung ist $\Delta p_2 = m_2(v_2' - v_2) = 2\mu(v_1 - v_2)$. Bei gleichen Massen $m_1 = m_2$ werden einfach die Geschwindigkeiten vertauscht. Für den Spezialfall, daß m_2 vor dem Stoß ruht ($v_2 = 0$), sind v_2' und E_2' in Abhängigkeit von m_1/m_2 in Abb. 2.1.10–1 dargestellt. Mit steigendem Verhältnis m_1/m_2 kann v_2' bis auf $2v_1$ anwachsen. Die auf m_2 übertragene Energie $E_2' = 4m_1m_2(m_1 + m_2)^{-2}E_1$ ist für $m_1 = m_2$ am größten. Wenn man also Teilchen durch Stöße abbremsen will, dann muß man als Stoßpartner etwa gleich schwere Teilchen anbieten. Zum Abbremsen von Neutronen ist daher Wasserstoff als Moderator am besten geeignet.

Unter den *inelastischen Stößen* soll nur der Fall betrachtet werden, bei dem die beiden Massen nach dem Stoß aneinander haften bleiben oder ein Geschoß m_1 in dem Körper m_2 stecken bleibt, also $v_1' = v_2'$ gilt. Aus der Impulsbilanz folgt $v' = P/M = (m_1v_1 + m_2v_2)/(m_1 + m_2)$ und aus der Energiebilanz $\alpha = (P^2/2M)/E$. Nach dem inelastischen Stoß besteht die mechanische Energie allein aus der mit der Schwerpunktbewegung verknüpften kinetischen Energie.

2.1.10.3. Elastischer Stoß zweier Massen. Beim räumlichen elastischen Stoß ist der Bewegungsablauf durch die Erhaltungsgrößen weitgehend festgelegt. Der Schwerpunkt bewegt sich mit konstanter Geschwindigkeit auf einer Geraden. Im Schwerpunktsystem erfolgt die Bewegung in einer Ebene, in der der Schwerpunkt liegt und zu der der Drehimpuls Normalenvektor ist.

Der Zusammenhang der Impulse $\boldsymbol{p}_1$ und $\boldsymbol{p}_2$ vor dem Stoß mit dem Schwerpunktimpuls $\boldsymbol{P}$ und dem Impuls im Schwerpunktsystem $\boldsymbol{p}$ ist in Abb. 2.1.10–2 veranschaulicht. Da wegen der Energieerhaltung beim elastischen Stoß der Betrag des Impulses im Schwerpunktsystem vor und nach dem Stoß gleich groß ist, liegen alle möglichen Vektoren $\boldsymbol{p}'$ auf einem Kreis. Sie können durch den Ablenkwinkel ϑ charakterisiert werden.

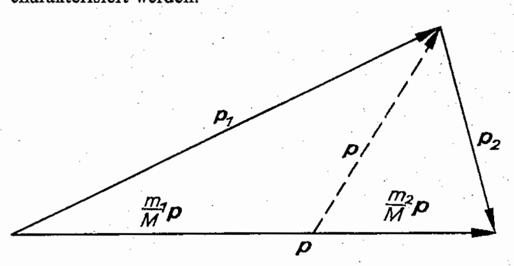


Abb. 2.1.10–2 Impulse im Labor- und Schwerpunktsystem

Für den Fall, daß m_2 vor dem Stoß ruht, $\boldsymbol{p}_2 = 0$, ist der Stoß in Abb. 2.1.10–3 dargestellt. Man sieht, daß der Ablenkwinkel im Laborsystem ϑ_L kleiner als ϑ ist. Bei gleichen Massen $m_1 = m_2$ gilt $\vartheta_L = \vartheta/2$. Für $m_1 < m_2$ kann der Ablenkwinkel alle Werte $\vartheta_L = 0, \ldots, \pi$ annehmen, für $m_1 > m_2$ gibt es einen maximalen Ablenkwinkel und zwei Stoßparameter, die zu dem gleichen Ablenkwinkel führen.

2.1.10.4. Wirkungsquerschnitt. Bei einem Streuexperiment (Abb. 2.1.10–4) beobachtet man z. B. mit einem Zählrohr den Teilchenstrom, der in einen bestimmten, durch die Öffnung des Zählrohres gegebenen Raumwinkel abgelenkt wird. Die charakteristische Meßgröße ist der Wirkungsquerschnitt.

Wie der Stoß verläuft, ist durch den Stoßparameter b (Abb. 2.1.10–5) und das Kraftfeld bzw. Potential $V(r)$ bestimmt. Durch Lösen der Bewegungsgleichung erhält man den Zusammenhang $b = b(\vartheta)$.

Der einfallende Teilchenstrom sei durch eine Stromdichte j_0 gekennzeichnet. Durch eine Fläche $\mathrm{d}A$ (vgl. Abb. 2.1.10–4) treten dann $\mathrm{d}N_0 = j_0\mathrm{d}A\,\mathrm{d}t$ Teilchen in der Zeit $\mathrm{d}t$ hindurch. Ist die Stromdichte der gestreuten Teilchen $j(r, \vartheta)$, wobei wegen der Teilchenzahlerhaltung $j \sim 1/r^2$ gilt, dann treten in das registrierende Meßinstrument $j(r, \vartheta)r^2\,\mathrm{d}\Omega\,\mathrm{d}t$ Teilchen in der Zeit $\mathrm{d}t$ ein (vgl. Abb. 2.1.10–5). Da j proportional zur Stromdichte j_0 der einfallenden Teilchen ist, setzt man $j(r, \vartheta)r^2 = j_0\sigma(\vartheta)$. $\sigma(\vartheta)$ wird als *differentieller Wirkungsquerschnitt* bezeichnet. Im nichtkugelsymmetrischen Potential kann er auch von φ abhängen. Als Integral über alle Raumwinkel erhält man den *Gesamtwirkungsquerschnitt* $\sigma = \int \mathrm{d}\Omega\,\sigma(\vartheta, \varphi)$. Wählt man $\mathrm{d}A$ (vgl. Abb. 2.1.10–4) von der Größe σ, dann wird $\mathrm{d}N_0$ gleich

Abb. 2.1.10–3 Stoß mit $p_2 = 0$, (a) bei $m_1 < m_2$ und (b) bei $m_1 > m_2$

der Summe aller gestreuten Teilchen. Der Streuer wirkt also wie ein Hindernis von der Querschnittsfläche σ.

Man betrachtet Bahnen (vgl. Abb. 2.1.10–5), deren Stoßparameter um db variieren, also dA = $b\,\mathrm{d}\varphi\,\mathrm{d}b$. Da aus der Bewegungsgleichung ein eindeutiger, stetiger Zusammenhang $b(\vartheta)$ folgt, liegen diese Bahnen nach dem Stoß im Raumwinkel $\mathrm{d}\Omega = \sin\vartheta\,\mathrm{d}\vartheta\,\mathrm{d}\varphi$, also gilt $\mathrm{d}N = \mathrm{d}N_0$. Der Vergleich von $\mathrm{d}N_0 = j_0 b\,\mathrm{d}\varphi\,\mathrm{d}b\,\mathrm{d}t$ und $\mathrm{d}N = j_0\sigma(\vartheta)\mathrm{d}\Omega\,\mathrm{d}t$ liefert für den differentiellen Wirkungsquerschnitt – unter Beachtung, daß ϑ mit wachsendem b kleiner wird –

$$\sigma = -\frac{b}{\sin\vartheta}\,\frac{\mathrm{d}b}{\mathrm{d}\vartheta}.$$

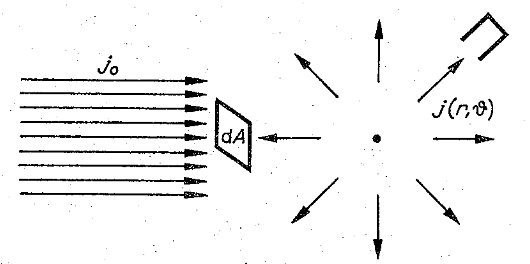


Abb. 2.1.10–4 Streuexperiment; einfallender Teilchenstrom, Streuzentrum und Zählrohr

Abb. 2.1.10–5 Stoßparameter und Ablenkwinkel

Abb. 2.1.10–6 Streuung an einer harten Kugel

Für eine starre, reflektierende Kugel (Abb. 2.1.10–6) ist $b = R\cos\vartheta/2$. Daraus ergibt sich der differentielle Wirkungsquerschnitt $\sigma(\vartheta) = R^2/4$ und der Gesamtwirkungsquerschnitt πR^2. Dieser hat den Wert des Kugelquerschnittes.

2.1.10.5. Rutherford-Streuung. Die Streuung von α-Teilchen an Atomkernen ist historisch interessant, weil derartige Experimente RUTHERFORD 1911 zu dem Schluß führten, daß die positive Ladung und die Hauptmasse eines Atoms in einem relativ kleinen Raumbereich des Atoms, dem Atomkern, lokalisiert sein müssen.

Der Streuung von Ladungen im Coulomb-Potential $V(r) = qQ/4\pi\varepsilon_0 r$ entsprechen die Streuzustände der in 2.1.6.4. besprochenen Bewegung im Potential $-C/r$. Bei Berücksichtigung der Mitbewegung des Kernes ist nur m durch die reduzierte Masse zu ersetzen. r ist die Differenzkoordinate. Um den Wirkungsquerschnitt ausrechnen zu können, wird der Zusammenhang zwischen dem Stoßparameter b und dem Ablenkwinkel ϑ benötigt.

Die in Abb. 2.1.10–7 eingetragenen beiden Hyperbelbahnen gehören bei gleichen Werten von E und L zu unterschiedlichem Vorzeichen von C. Der Ablenkwinkel ϑ ist jedoch der gleiche.

Ist die Geschwindigkeit für das weit entfernte Teilchen v, dann gilt $W = mv^2/2$, $L = mvb = \sqrt{2mW}\,b$. Statt des Drehimpulses wird hier also der Stoßparameter zur Unterscheidung verschiedener Bahnen verwendet. Aus $\cos\varphi_\infty = -1/\varepsilon$, $\varphi_\infty = \pi/2 + \vartheta/2$ und $\varepsilon^2 - 1 = \cot^2\vartheta/2$ sowie L^2

$= m(C/2W)^2(\varepsilon^2 - 1)$ folgt $b(\vartheta) = (|C|/2E)\cot\vartheta/2$. Dies führt auf den Wirkungsquerschnitt

$\sigma(\vartheta) = \left(\frac{qQ}{4\pi\varepsilon_0}\frac{1}{4W}\right)^2 \frac{1}{\sin^4\vartheta/2}$, auch als Rutherfordsche Streuformel bezeichnet. Sie gibt den Wirkungsquerschnitt im Schwerpunktsystem an, der bei $m_2 \gg m_1$ praktisch mit dem im Laborsystem übereinstimmt, sonst muß die Umrechnung auf das Laborsystem erfolgen.

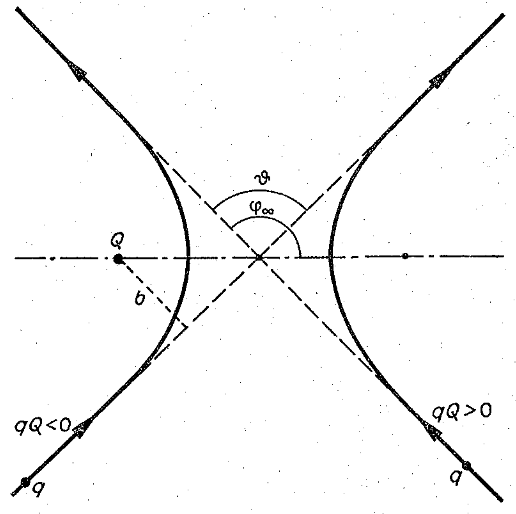


Abb. 2.1.10–7 Streuung im Coulomb-Feld

2.1.11. Starre Körper

Der starre Körper ist ein Modell eines Festkörpers, in dem man von Deformationen absieht. Die einzelnen Massenelemente des Festkörpers werden als starr miteinander verknüpft betrachtet, die Abstände und die Winkel zwischen verschiedenen Verbindungen sind unveränderlich. Von einem System von Massenpunkten ausgehend kann man diese starren Verbindungen als Zwangsbedingungen auffassen, die die Zahl der Freiheitsgrade vermindern.

Der starre Körper besitzt sechs *Freiheitsgrade*. Drei Zahlenangaben benötigt man, um die Lage eines Punktes des Festkörpers, z. B. seines Schwerpunktes, anzugeben. Bei festgehaltenem Schwerpunkt kann eine durch den Schwerpunkt laufende Achse beliebig im Raum orientiert sein. Ihre Richtung e_3 ist durch zwei Winkel ϑ und φ bestimmt, wie ein Punkt auf einer Kugeloberfläche durch Breiten- und Längenkreis (Abb. 2.1.11–1). Werden Schwerpunkt und Orientierung einer Achse festgehalten, dann kann man die Lage des Festkörpers nur noch dadurch verändern, daß man ihn um diese Achse dreht. Diese Drehung kann durch einen weiteren Winkel χ angegeben werden. Man nennt diese drei Winkel die *Eulerschen Winkel*. Sie werden von den in Abb. 2.1.11-1 eingezeichneten Achsen aus gezählt.

Schwerpunktbewegung. Die Schwerpunktbewegung $M\ddot{s} = F$ wird durch die auf den Festkörper einwirkende Gesamtkraft F bestimmt. Verschwindet diese, dann ist der Gesamtimpuls P eine Erhaltungsgröße und der Schwerpunkt bewegt sich auf der Geraden $s = (P/M)t + s_0$.

Drehbewegung. Die Drehbewegung folgt aus der Bewegungsgleichung $\dot{L} = M$. Wirkt kein Drehmoment M, dann ist der Drehimpuls L (bezüglich des Schwerpunktes) zeitlich konstant.

2.1.11.1. Trägheitsmoment. Bei einer Drehung um eine Achse mit einer Winkelgeschwindigkeit ω (Abb. 2.1.11–2) hat ein Massenelement dm im Abstand a von der Drehachse die Geschwindigkeit $a\omega$. Seine kinetische Energie ist

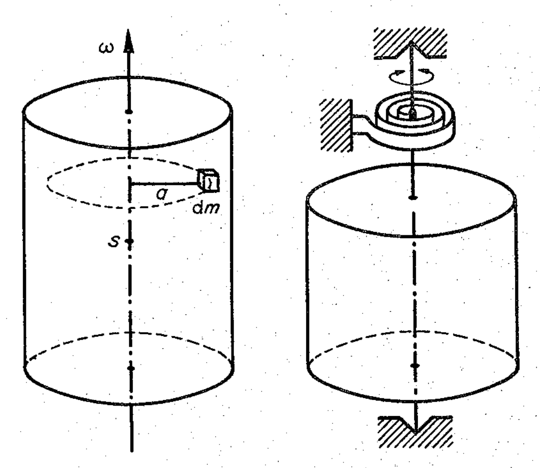


Abb. 2.1.11–2 Kinetische Energie eines Massenelementes
Abb. 2.1.11–4 Drehschwingung (*rechts*)

daher $dW_{kin} = (1/2)dma^2\omega^2$. Da alle Massenelemente mit der gleichen Winkelgeschwindigkeit rotieren, ist die gesamte kinetische Energie $W_{kin} = (1/2)\Theta_A\omega^2$ mit $\Theta_A = \int dm\, a^2$. Θ_A ist das Massenträgheitsmoment bezüglich der Achse A. Für regelmäßig geformte Körper kann die Integration analytisch ausgewertet werden, einige Trägheitsmomente sind in Tab. 2.1.11–3 zusammengestellt. Sie beziehen sich auf Achsen, die durch den Schwerpunkt gehen. Hat der Schwerpunkt eine Entfernung s von der Drehachse, dann ist mit der Drehung eine Schwerpunktbewegung verknüpft, deren kinetische Energie $(1/2)Ms^2\omega^2$ beträgt. Die Trägheitsmomente bezüglich einer Achse durch den Schwerpunkt Θ_S und bezüglich einer Achse gleicher Richtung im Abstand s vom Schwerpunkt Θ_A sind über $\Theta_A = \Theta_S + Ms^2$ (*Steinerscher Satz*) miteinander verknüpft. Die Drehimpulskomponente des Massenelementes dm (vgl. Abb. 2.1.11-2) in Richtung der Achse A ist $dL_A = dm\, a^2\omega$, also $L_A = \Theta_A\omega$.

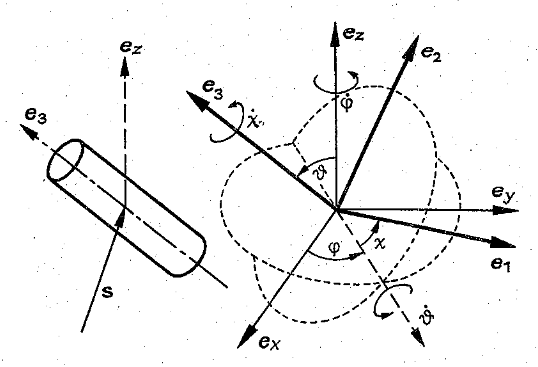


Abb. 2.1.11–1 Freiheitsgrade des starren Körpers

Tab. 2.1.11–3 Trägheitsmomente

Körper	Formel
Vollkugel	$\Theta = \frac{2}{5} MR^2$
Hohlkugel	$\Theta = \frac{2}{3} MR^2$
Vollzylinder	$\Theta_3 = \frac{1}{2} MR^2, \; \Theta_1 = \frac{1}{4} M \, (R^2 + h^2/3)$
Hohlzylinder	$\Theta_3 = MR^2$
Quader	$\Theta_3 = \frac{1}{12} M \, (a^2 + b^2)$
Stab	$\Theta_1 = \frac{1}{3} Ml^2$

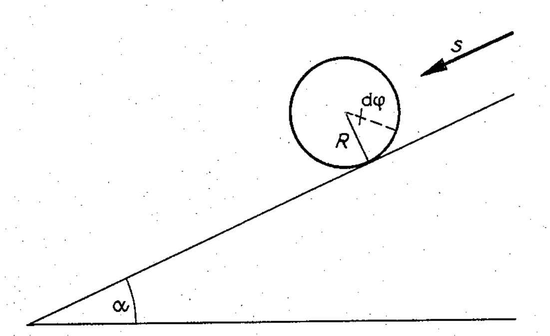


Abb. 2.1.11–5 Zylinder auf einer schiefen Ebene

Das Trägheitsmoment ist die die Massenverteilung des Festkörpers charakterisierende Größe. Sie wird zur Berechnung physikalischer Größen wie kinetische Energie und Drehimpuls benötigt.

2.1.11.2. Drehung um eine Achse. Bei Drehungen um eine Symmetrieachse treten nicht die beim Kreisel zu beschreibenden Komplikationen auf. Die Bewegungsgleichung lautet $\dot{L} = \dot{M}$, $L = \Theta\omega$.

Drehschwingungen. Bei ihnen tritt ein rücktreibendes Drehmoment proportional zum Auslenkungswinkel φ auf, $\dot{M} = -D\varphi$ (Abb. 2.1.11–4). Das Richtmoment wird mit D bezeichnet. Mit $\omega = \dot{\varphi}$ lautet die Bewegungsgleichung $\Theta\ddot{\varphi} + D\varphi = 0$, die Lösungen sind Schwingungen mit der Kreisfrequenz $\omega = \sqrt{D/\Theta}$ bzw. der Schwingungsdauer $T = 2\pi\sqrt{\Theta/D}$.

Ein auf einer schiefen Ebene *abrollender Zylinder* (Abb. 2.1.11–5) hat einen Freiheitsgrad. Bewegung des Schwerpunktes und Drehung um die Zylinderachse sind durch $\dot{\varphi} = \dot{s}/R$ verknüpft. Die kinetische Energie des Zylinders ist daher $T = (1/2)M\dot{s}^2 + (1/2)\Theta(\dot{s}/R)^2$, die potentielle Energie im Schwerefeld $V = -Mgs\sin\alpha$ ist durch die Höhe des Schwerpunktes bestimmt. Analog zur abgleitenden Masse (vgl. 2.1.4.3.) ist die Bewegungsgleichung $(M + \Theta/R^2)\ddot{s} = Mg\sin\alpha$. Die Beschleunigung des Schwerpunktes wird dadurch vermindert, daß ein Teil der potentiellen Energie in zur Rotation des Zylinders gehörige kinetische Energie umgewandelt wird, während beim Abgleiten die gesamte potentielle Energie zur Erhöhung der kinetischen Energie des Schwerpunktes zur Verfügung steht.

Zylinder gleicher Radien und gleicher Masse können verschiedene Trägheitsmomente haben (vgl. Tab. 2.1.11–3). Ein Vollzylinder rollt daher schneller ab als ein Hohlzylinder.

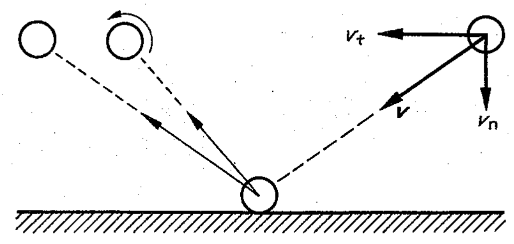


Abb. 2.1.11–6 Reflexion eines Balles

In Abb. 2.1.11–6 wird der Aufprall eines Balles auf eine Unterlage als elastischer Stoß betrachtet. Die Geschwindigkeit des ankommenden Balles wird in eine tangentiale v_t und eine normale Komponente v_n zerlegt. Die Normalkomponente hat nach dem elastischen Stoß ihre Richtung umgekehrt. Das Verhalten der Tangentialkomponente wird dadurch bestimmt, ob der Ball während der Zeit der Berührung gleitet oder haftet. Im ersten Fall ändert sich v_t nicht, im zweiten Fall wird der Ball in Drehung versetzt. Dabei wird ein Teil der mit der Bewegung v_t verknüpften kinetischen Energie in Rotationsenergie umgewandelt, die Tangentialkomponente der Geschwindigkeit verringert sich also um einen Faktor $(1 + \Theta/MR^2)^{-1/2} \approx 0{,}77$. Diesen Unterschied muß man vor Augen haben, wenn beim Fußball davon gesprochen wird, daß „der Ball auf nassem Rasen schnell wird".

2.1.11.3. Physikalisches Pendel. Bei der Schwingung eines ausgedehnten Körpers um eine horizontale Achse (Abb. 2.1.11–7) tritt bei kleinen Auslenkungen ein rücktreibendes Drehmoment $\dot{M} = -Mgs\varphi$ auf, wobei s der Abstand des Schwerpunktes von der Achse ist. Die Kreisfrequenz ist daher $\omega = \sqrt{Mgs/\Theta_A} = \sqrt{g/l^*}$. Als reduzierte Pendellänge bezeichnet man $l^* = s(1 + \Theta/Ms^2)$. Ein mathematisches Pendel mit der Fadenlänge $l = l^*$ hat die gleiche Schwingungsdauer. Legt man die Achse durch A′ (vgl. Abb. 2.1.11–7) mit einem Abstand $l^* - s$ vom Schwerpunkt, dann erhält man die gleiche Schwingungsdauer.

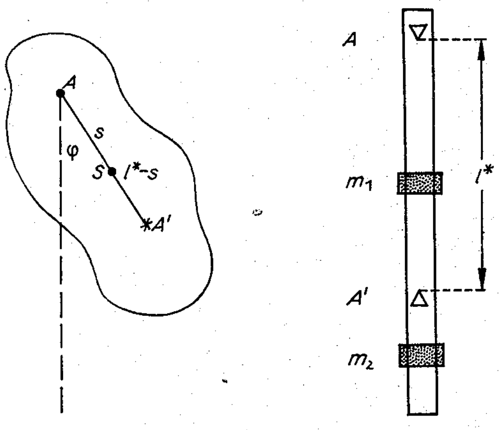


Abb. 2.1.11–7 Physikalisches Pendel (*links*)
Abb. 2.1.11–8 Reversionspendel

Dieser Sachverhalt wird z. B. beim *Reversionspendel* (Abb. 2.1.11-8) ausgenutzt. Man mißt den Abstand zweier Achsen A und A' auf einem Pendelstab genau aus. Verschiebbare Massen werden auf dem Stab dann so verlagert, daß die Schwingungsdauer um beide Achsen gleich ist. Durch Beobachten genügend vieler Schwingungen ist diese Schwingungsdauer auch genau meßbar. Auf diesem Wege kann man den lokalen Wert der Erdbeschleunigung mit hoher Genauigkeit messen.

2.1.11.4. Bewegungsgleichung. Die Bewegungsgleichung des starren Körpers kann aufgestellt werden, wenn die Lagrange-Funktion bekannt ist. Ist $\boldsymbol{r}_n$ die Position eines Massenelementes m_n in bezug auf den Schwerpunkt S (Abb. 2.1.11-9), dann ist sein Ortsvektor durch $\boldsymbol{s} + \boldsymbol{r}_n$ gegeben. Seine Geschwindigkeit beträgt $\dot{\boldsymbol{s}} + \boldsymbol{\omega} \times \boldsymbol{r}_n$, wobei $\boldsymbol{\omega}$ der Vektor der *momentanen Winkelgeschwindigkeit* ist, die die

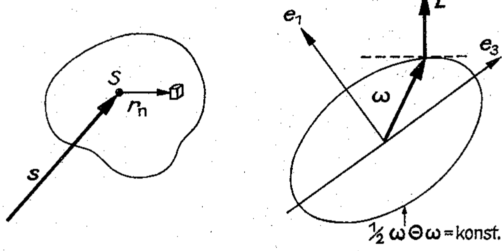


Abb. 2.1.11-9 Position eines Massenelementes
Abb. 2.1.11-10 Trägheitsellipsoid (*rechts*)

Drehung um den Schwerpunkt beschreibt. Damit wird die kinetische Energie $T = \sum_n (1/2) m_n (\dot{\boldsymbol{s}} + \boldsymbol{\omega} \times \boldsymbol{r}_n)^2 = (1/2) M \dot{s}^2 + (1/2) \boldsymbol{\omega} \Theta \boldsymbol{\omega}$. Θ ist der *Trägheitstensor*

$$\Theta = \sum_n m_n [r_n^2 I - \boldsymbol{r}_n \circ \boldsymbol{r}_n] = \int \mathrm{d}m [r^2 I - \boldsymbol{r} \circ \boldsymbol{r}],$$

$$\Theta_{ik} = \int \mathrm{d}m [r^2 \delta_{ik} - x_i x_k].$$

Er ist symmetrisch und hat daher drei senkrecht zueinander stehende Hauptachsen $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$. Falls der Körper Symmetrieachsen hat, fallen die Hauptachsen in deren Richtung. Der Drehimpuls $\boldsymbol{L} = \sum_n (\boldsymbol{s} + \boldsymbol{r}_n) \times m_n (\dot{\boldsymbol{s}} + \boldsymbol{\omega} \times \boldsymbol{r}_n)$ vereinfacht sich zu $\boldsymbol{L}_{\text{ges}} = \boldsymbol{s} \times M \dot{\boldsymbol{s}} + \Theta \boldsymbol{\omega} = \boldsymbol{L}_S + \boldsymbol{L}$. Kinetische Energie und Drehimpuls spalten in zwei Summanden auf, die den mit der Schwerpunktbewegung bzw. mit der Rotation um den Schwerpunkt verknüpften Anteil beschreiben.

$\boldsymbol{L} = \Theta \boldsymbol{\omega}$ drückt aus, daß der Trägheitstensor jedem Vektor der Winkelgeschwindigkeit einen Vektor Drehimpuls zuordnet. Die Richtung beider Vektoren ist i. allg. verschieden, nur wenn $\boldsymbol{\omega}$ in Richtung einer Hauptachse des Trägheitstensors zeigt, liegen beide parallel. Betrachtet man als Funktion von $\boldsymbol{\omega}$ eine Fläche konstanter kinetischer Energie $T = (1/2) \boldsymbol{\omega} \Theta \boldsymbol{\omega} = \text{konst.}$ (Abb. 2.1.11-10), dann gibt die Normale an dieses Ellipsoid die zu $\boldsymbol{\omega}$ gehörige Richtung des Drehimpulses an.

Die potentielle Energie ist $V = \sum_n v(\boldsymbol{s} + \boldsymbol{r}_n)$. Damit ist die Lagrange-Funktion bekannt. Aus ihr erhält man die Bewegungsgleichungen

$$m \ddot{\boldsymbol{s}} = \boldsymbol{F}, \quad \boldsymbol{F} = \sum_n \boldsymbol{F}_n \quad \text{bzw.} \quad \dot{\boldsymbol{L}} = \boldsymbol{\check{M}}, \quad \boldsymbol{\check{M}} = \sum_n \boldsymbol{r}_n \times \boldsymbol{F}_n$$

für die Schwerpunktbewegung bzw. für die Drehbewegung um den Schwerpunkt. Man beachte, daß hierin $\boldsymbol{L}$ bzw. $\boldsymbol{\check{M}}$ der Drehimpuls bzw. das Drehmoment bezüglich des Schwerpunktes sind.

2.1.11.5. Kreisel. Wirkt auf einen starren Körper kein Drehmoment, dann ist der Drehimpuls eine Erhaltungsgröße. Dies ist im Schwerefeld z.B. dann der Fall, wenn der Körper in seinem Schwerpunkt unterstützt wird. Man spricht dann von einem *kräftefreien Kreisel*.

Zur Vereinfachung werden nur Körper mit einer Symmetrieachse ($\boldsymbol{e}_3$) betrachtet. Für diese gilt $\Theta_1 = \Theta_2$. Aus Abb. 2.1.11-10 ist zu ersehen, daß die drei Vektoren $\boldsymbol{L}$, $\boldsymbol{e}_3$ und $\boldsymbol{\omega}$ in einer Ebene ligen. Wegen $\Theta_1 = \Theta_2$ kann $\boldsymbol{e}_1$ zu allen Zeiten als eine in dieser Ebene liegenden Richtung gewählt werden. Diese Ebene dreht sich im Zeitablauf um $\boldsymbol{L}$. Dabei beschreiben *Figurenachse F* ($\boldsymbol{e}_3$) und *momentane Drehachse* $\boldsymbol{\omega}$ Kegel um den konstanten Vektor des Drehimpulses, Nutationskegel bzw. Spurkegel genannt. Die Kreisfrequenz dieser *Nutation* ist $\omega_{\text{Nut.}} = L/\Theta_1$. Die Bewegung des starren Körpers kann man als Überlagerung der Nutation der Figurenachse um den Drehimpuls und einer Drehung um die Figurenachse beschreiben. Der Vektor der momentanen Winkelgeschwindigkeit beschreibt dabei – vom Körper aus gesehen – einen Kegel um die Figurenachse, den *Rastpolkegel*. Die Gesamtbewegung kann somit auch als Abrollen des Rastpolkegels auf dem Spurkegel erklärt werden. Für ein abgeplattetes bzw. verlängertes Trägheitsellipsoid wird dies in Abb. 2.1.11-11 gezeigt.

Schwerer symmetrischer Kreisel. Wie in Abb. 2.1.11-12 wird der Fall betrachtet, daß nicht der Schwerpunkt S, sondern ein Punkt P auf der Figurenachse im Abstand s vom

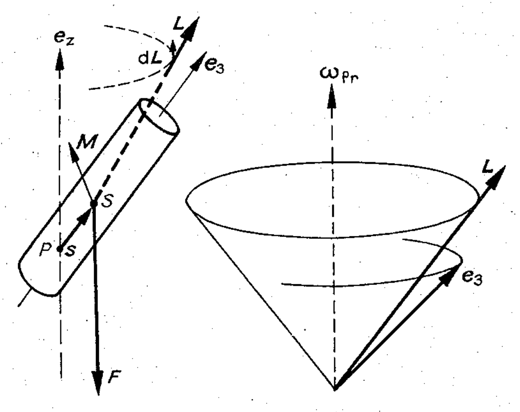
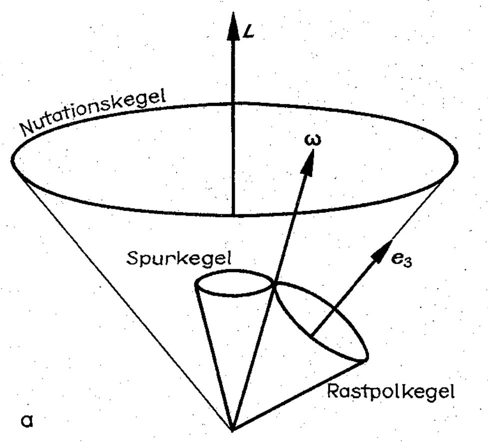


Abb. 2.1.11-12 Präzession

Schwerpunkt festgehalten wird. Die Bewegungsgleichung für die Drehbewegung um diesen Punkt lautet in diesem Fall auch $\dot{\boldsymbol{L}} = \boldsymbol{\check{M}}$, wobei $\boldsymbol{L}$ und $\boldsymbol{M}$ bezüglich des Punktes P zu berechnen sind. Es gilt wieder $\boldsymbol{L} = \Theta \boldsymbol{\omega}$, wenn man im Trägheitstensor $\Theta_1 = \Theta_2$ durch $\Theta_1 + Ms^2$ ersetzt. Es wirkt jetzt ein Drehmoment $\boldsymbol{\check{M}} = \boldsymbol{s} \times \boldsymbol{F}$ mit $\boldsymbol{F} = M\boldsymbol{g}$. Der Vektor Drehmoment $\boldsymbol{\check{M}}$ steht senkrecht auf $\boldsymbol{s}$, d.h. auf der Figurenachse, und senkrecht zu $\boldsymbol{F}$, d.h. er liegt horizontal. Damit sind die Komponenten L_z und L_3 Erhaltungsgrößen.

Ist zum Anfangszeitpunkt $L_2(0) = 0$, dann liegt $\boldsymbol{L}$ in der von der Figurenachse und der Vertikalen aufgespannten Ebene, dann steht also $\boldsymbol{\check{M}} \perp \boldsymbol{L}$. Aus der Bewegungsgleichung $\dot{\boldsymbol{L}} = \boldsymbol{\check{M}}$ folgt, daß durch das Drehmoment nur die Richtung, nicht der Betrag des Drehimpulses geändert wird.



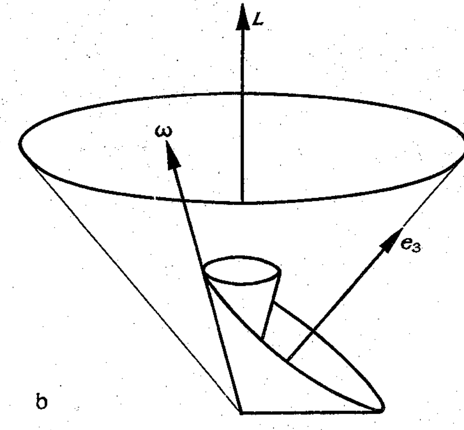


Abb. 2.1.11–11 Kräftefreier Kreisel, (a) bei $\Theta_1 > \Theta_3$ und (b) bei $\Theta_1 < \Theta_3$

Der Drehimpuls präzessiert mit der Winkelgeschwindigkeit $\omega_{\mathrm{Pr}} = (Mgs \sin \vartheta_{\mathrm{F}})/(L \sin \vartheta_{\mathrm{L}})$ (vgl. Abb. 2.1.11–12).
Ist $L_2 \neq 0$, dann führt die Figurenachse eine Nutation um die Drehimpulsrichtung aus. Für den Fall, daß das wirkende Drehmoment klein ist, erfolgt diese mit $\omega_{\mathrm{Nut}} = L/(\Theta_1 + Ms^2)$. Die *Präzessionsfrequenz* des Drehimpulses um die Vertikale ist in diesem Falle $\omega_{\mathrm{Pr}} = Mgs \cos \alpha / L$. α ist hierin der Winkel zwischen der Richtung von L und der Figurenachse. Der Durchstoßpunkt der Figurenachse beschreibt die in Abb. 2.1.11–13 dargestellten Kurven auf einer Kugeloberfläche.

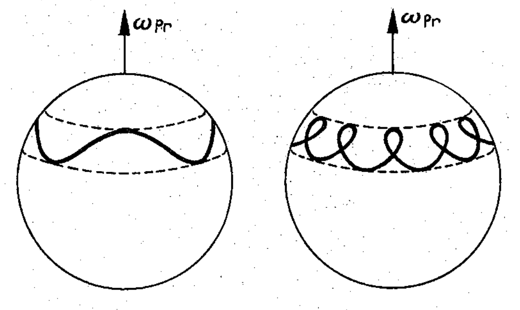


Abb. 2.1.11–13 Bahn der Figurenachse als mit Nutation überlagerte Präzession

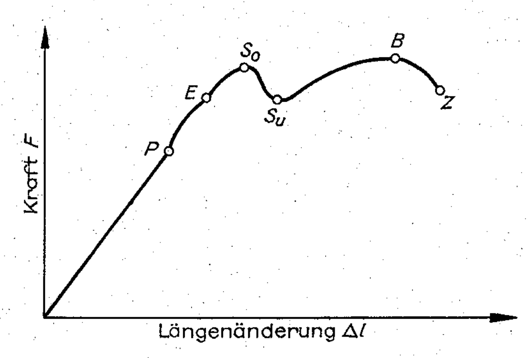


Abb. 2.2.1–1 Verformungskurve

2.2. Mechanik der Kontinua

2.2.1. Statisches elastisches Verhalten von Festkörpern

Die Beanspruchung eines festen Körpers durch eine äußere Kraft führt zu einer Deformation des Körpers. Das charakteristische Verhalten wird am einfachsten durch einen Zug- oder Druckversuch veranschaulicht. In Abb. 2.2.1–1 ist die Verformungskurve, d. h. die Abhängigkeit der Längenänderung Δl von der aufgewendeten Kraft F, schematisch dargestellt. Bis zur *Proportionalitätsgrenze P* hängen F und Δl linear miteinander zusammen. Oberhalb P ist die Verformungskurve gekrümmt, d. h. die Längenänderung bei Erhöhung der Kraft um einen festen Betrag ist größer als unterhalb der Proportionalitätsgrenze. Die *Elastizitätsgrenze E* ist dadurch bestimmt, daß ein Körper nach Entlastung den Ausgangszustand wieder erreicht, wenn er höchstens bis E belastet wurde. Oberhalb der Elastizitätsgrenze tritt bleibende *plastische Verformung* auf. Die weitere Belastung führt zur *oberen Streckgrenze* S_o. Nach einem Abfall der äußeren Kraft bis zur unteren Streckgrenze S_u ist zur weiteren Verformung ein erneuter Kraftanstieg notwendig. Im Punkt B wird die Kraft erreicht, die der Körper maximal ertragen kann. Hier setzt beim Zugversuch in einem begrenzten Abschnitt der Probe eine starke Querschnittsverminderung ein. Bei zunehmender Längenänderung nimmt die Kraft etwas ab und im Punkt Z zerreißt die Probe. Bei einigen Materialien, z. B. bei reinen Metallen, wird die *Streckgrenzenüberhöhung* nicht beobachtet, d. h., untere und obere Streckgrenze fallen in einem Punkt zusammen. Die Streckgrenze stimmt dann annähernd mit der Elastizitätsgrenze überein. Es ist in diesem Fall auch üblich, den Punkt, an dem plastische Verformung einsetzt, als *Fließgrenze* zu bezeichnen.
Die Plastizierung des Materials ist mit dem Auftreten einer *mechanischen Hysterese* verbunden (Abb. 2.2.1–2). Wird entlastet, nachdem zuvor auf der jungfräulichen Kurve *1* die Elastizitätsgrenze E überschritten worden war, so gibt es bei $F = 0$ eine bleibende Verformung (Entlastungskurve *2*), die nur durch eine entgegengesetzt wirkende Kraft zu Null gemacht werden kann. Wird weiter bis zu dem zum Punkt A symmetrisch gelegenen Punkt A' mit dieser zur ursprünglichen Kraft entgegen gerichteten belastet und dann wiederum die Kraftrichtung umgekehrt (Kurve *3*), so ergibt

sich die *Hysteresisschleife*. Deren Flächeninhalt stellt den bei der Verformung geleisteten Arbeitsbetrag dar, der bei diesem irreversiblen Prozeß in Wärme umgewandelt wird.

2.2.1.1. Spannung und Dehnung. Die Verformung eines Festkörpers kann sowohl durch *Massenkräfte*, die an jedem Volumenelement des Körpers angreifen – als Beispiel sei die Schwerkraft genannt –, als auch durch *Oberflächenkräfte*, die an einzelnen Oberflächenelementen wirken, verursacht werden. Wegen ihrer Bedeutung für die praktischen Anwendungen sollen im folgenden ausschließlich die Wirkungen der Oberflächenkräfte betrachtet werden.

Die Größe der Deformation, die eine äußere Kraft im Festkörper hervorruft, hängt nicht nur vom absoluten Betrag der Kraft ab, sondern auch von der Größe der Fläche, über die die Kraft verteilt ist. Dadurch erklärt sich z.B. die Tatsache, daß ein Fußgänger im Schnee wesentlich tiefer einsinkt, als ein Skiläufer, da bei gleichbleibender Kraft (Gewicht der Person) im zweiten Fall die Fläche ein Vielfaches der im ersten ist. Es wird deshalb die *mechanische Spannung* σ als Quotient aus Kraft F und Fläche A als die für die Deformation entscheidende Größe eingeführt, $\sigma = F/A$.

Die äußeren Spannungen rufen als Reaktion im Körper innere Spannungen hervor. Wird ein beliebiges Volumenelement im Inneren des Körpers betrachtet (Abb. 2.2.1–3), so greift die Spannung an jeder Fläche an. Da sie aber i. allg. nicht senkrecht auf der herausgegriffenen Fläche steht, wird der allgemeine Spannungszustand durch drei Spannungskomponenten charakterisiert. In Abb. 2.2.1–3 sind für eine Begrenzungsfläche x = konst. des als würfelförmig angenommenen Volumenelementes die drei Spannungskomponenten eingezeichnet. Dabei ist die *Normalspannung* σ_{xx} von den beiden *Tangentialspannungen* τ_{xy} und τ_{xz} zu unterscheiden. Der erste Index gibt jeweils die Ebene an (hier x = konst.), auf die die Spannung bezogen ist, der zweite Index bezeichnet die Richtung, in die die entsprechende Spannungskomponente wirkt. Die Normalspannung wird auch als *Druck* oder *Druckspannung* bezeichnet und ist gleich der Normalkraft je Flächeneinheit. Bei negativem

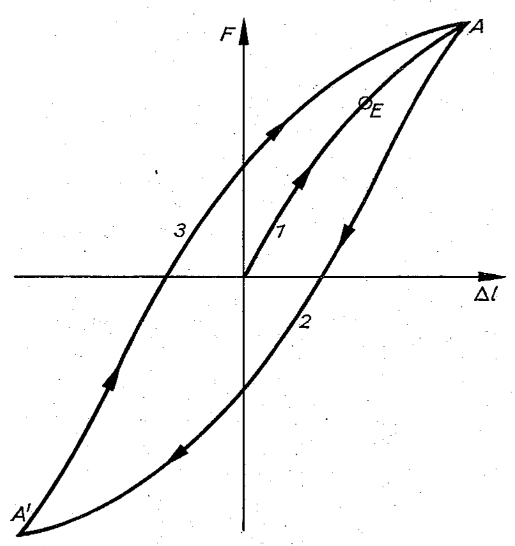


Abb. 2.2.1–2 Mechanische Hysterese

Vorzeichen handelt es sich um *Zug* bzw. um eine *Zugspannung*. Die Tangentialspannung oder *Schubspannung* ist gleich der entsprechenden Tangentialkraftkomponente pro Flächeneinheit. Da zu jeder Ebenenschar x = konst., y = konst. und z = konst. je drei Spannungskomponenten gehören, ist der vollständige Spannungszustand eines Volumenelementes durch neun Spannungskomponenten bestimmt, von denen wegen gewisser Symmetriebeziehungen bei den Schubspannungen zur Gewährleistung der Drehmomentenfreiheit, z.B. $\tau_{xy} = \tau_{yx}$, aber nur sechs voneinander unabhängig sind. Die Schubspannungen bilden die Nichtdiagonalkomponenten des Spannungstensors σ. Die durch die Spannungen hervorgerufene Verformung des Körpers

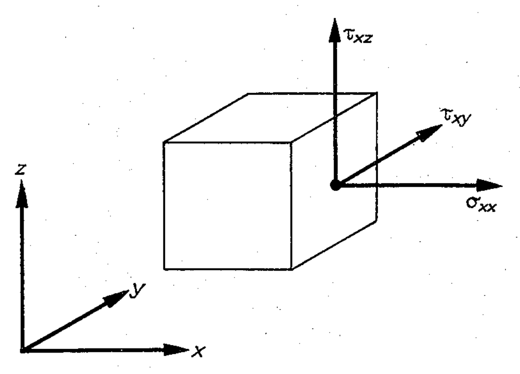


Abb. 2.2.1–3 Spannungskomponenten an der Fläche x = konst. eines Volumenelementes

wird durch die *Verzerrungen* beschrieben, wobei *Dehnungen* ε und *Schiebungen* γ unterschieden werden. Die Dehnungen sind die durch die entsprechenden Normalspannungen σ hervorgerufenen relativen Längenänderungen $\varepsilon = \Delta l/l$, wenn Δl die absolute Längenänderung und l die ursprüngliche Länge sind. Die Schiebungen sind die durch die Tangentialspannungen τ erzeugten Verzerrungswinkel im Bogenmaß (vgl. 2.2.1.2.). Da jeder Spannungskomponente eine Verzerrungskomponente ε bzw. γ entspricht, gibt es sechs unabhängige Verzerrungen. Statt der Verzerrungswinkel werden auch die Nichtdiagonalelemente des Verzerrungstensors $\varepsilon_{xy} = \gamma_{xy}/2$ usw. verwendet.

Der quantitative Zusammenhang zwischen Spannungen und Verzerrungen ist eine materialspezifische Eigenschaft des betrachteten Festkörpers, d.h., er muß experimentell ermittelt werden. Das gilt sowohl für die Art des funktionalen Zusammenhanges als auch für die Parameter, die in die entsprechenden Beziehungen eingehen. HOOKE zeigte, daß für hinreichend kleine Verformungen für alle Festkörper ein linearer Zusammenhang zwischen Spannung und Verzerrung vorhanden ist (vgl. den Verlauf der Verformungskurve in Abb. 2.2.1–1 bis zur Proportionalitätsgrenze P). Diese Aussage gilt allgemein für Reaktionen des Festkörpers auf einen schwachen äußeren Einfluß, hier einer kleinen Spannung. Da bei verschwindender Spannung keine Verzerrung auftritt und bei Umkehr des Vorzeichens der Spannung, d. h. Druck statt Zug, auch die Verzerrung ihr Vorzeichen wechselt, können im Zusammenhang zwischen Spannung und Verzerrung nur ungerade Potenzen auftreten, so daß es bei kleinen Verzerrungen ausreichend ist, den linearen Term zu berücksichtigen. Im allgemeinen muß davon ausgegangen werden, daß bei komplizierten

Spannungszuständen jede Spannungskomponente von jeder Verzerrungskomponente abhängig ist, so daß z. B. das *Hookesche Gesetz* in der allgemeinen Form mit den von VOIGT eingeführten Konstanten c_{ik}

$$\sigma_{xx} = c_{11}\varepsilon_{xx} + c_{12}\varepsilon_{yy} + c_{13}\varepsilon_{zz} + c_{14}\gamma_{yz} + c_{15}\gamma_{zx} + c_{16}\gamma_{xy},$$
$$\sigma_{yy} = c_{21}\varepsilon_{xx} + c_{22}\varepsilon_{yy} + c_{23}\varepsilon_{zz} + c_{24}\gamma_{yz} + c_{25}\gamma_{zx} + c_{26}\gamma_{xy}$$

lautet und entsprechend für alle anderen Komponenten. Für größere Verformungen wird der Zusammenhang nichtlinear. Er ist i. allg. auch zeitabhängig (vgl. 2.2.3.2.).
Die klassische Elastizitätstheorie setzt den linearen Zusammenhang zwischen Spannung und Verzerrung voraus, so daß bei der Anwendung der Ergebnisse immer der Gültigkeitsbereich des Hookeschen Gesetzes zu beachten ist. Der große Vorteil dieser Beschränkung liegt darin, daß bei linearen Beziehungen das *Superpositionsprinzip* angewendet werden kann, d. h., verschiedene Spannungszustände überlagern sich, ohne sich gegenseitig zu beeinflussen. Dadurch gelingt es, auch kompliziertere Spannungszustände theoretisch zu behandeln.

2.2.1.2. Isotrope Medien. Bei einem isotropen Medium sind die Materialeigenschaften von der Beanspruchungsrichtung unabhängig, z. B. ist die Dehnung bei vorgegebener Normalspannung in jeder Raumrichtung gleich groß. Die meisten Materialien, die als Konstruktionswerkstoff eingesetzt werden, verhalten sich mechanisch isotrop. Die üblichen Baustähle z. B. haben ein polykristallines Gefüge, so daß keine Vorzugsrichtung vorhanden ist; dasselbe gilt für alle anderen Konstruktionsteile aus Metallen bzw. Metallegierungen, die im Leichtbau verwendet werden. Ziegelmauerwerk und Beton in den verschiedensten Varianten sind infolge ihrer Struktur ebenfalls mechanisch isotrop. Selbst Werkstoffe, die ihrer Natur nach eine Vorzugsrichtung besitzen (Wachstumsrichtung bei Holz, Verstrekkungsrichtung bei einigen Polymeren, Richtung der Stahleinlage bei Stahlbeton), werden näherungsweise wie mechanisch isotrope Materialien behandelt, wenn die Richtung der Hauptbeanspruchung mit dieser Vorzugsrichtung übereinstimmt. Zur Charakterisierung des elastischen Verhaltens eines isotropen Mediums reichen zwei Materialkonstanten aus. Werden dafür die beiden *Laméschen Konstanten* μ und λ benutzt, so lautet das Hookesche Gesetz für isotrope Medien

$$\sigma_{xx} = 2\mu\varepsilon_{xx} + \lambda(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})$$
$$\tau_{xy} = \mu\gamma_{xy} = 2\mu\varepsilon_{xy}$$

und analog für die anderen Normalspannungs- und Schubspannungskomponenten. Die Bedeutung der Konstanten c_{ik} in der allgemeinen Formulierung des Hookeschen Gesetzes (vgl. 2.2.1.1.) für den Fall isotroper Medien läßt sich durch Vergleich der entsprechenden Gleichungen sofort ablesen. Im folgenden wird die Wirkung einiger einfacher, in der Praxis häufig auftretender Beanspruchungsarten behandelt.

Einachsiger Zug oder Druck. Ein Stab oder Draht wird in Längsrichtung, die gleichzeitig die z-Richtung sei, belastet (Abb. 2.2.1–4). Handelt es sich um eine Druckspannung σ_{zz}, so ist die Verkürzung $-\Delta l$ des Stabes mit einer Vergrößerung der Querdimension ($+\Delta a$ in x-Richtung; $+\Delta b$ in y-Richtung; durch das Pluszeichen soll die Vergrößerung zum Ausdruck gebracht werden) verbunden. Die *Poissonsche Querkontraktionszahl* ν ist gleich dem Verhältnis der relativen Änderung der Querdimensionen zur relativen Längenänderung, $\nu = -(\Delta a/a)/(\Delta l/l) = -(\Delta b/b)/(\Delta l/l)$. Es sind die Dehnung in z-Richtung $\varepsilon_{zz} = \Delta l/l$ und die entsprechende in y-Richtung $\varepsilon_{yy} = \Delta b/b = -\nu\varepsilon_{zz}$ und in x-Richtung $\varepsilon_{xx} = \Delta a/a = -\nu\varepsilon_{zz}$. Da bei reinem Zug in z-Richtung die beiden Normalspannungen $\sigma_{xx} = \sigma_{yy} = 0$ sind, folgt aus dem Hookeschen Gesetz für diese Spannungskomponenten $2\nu = \lambda/(\mu + \lambda)$. Damit wird die Spannung

$$\sigma_{zz} = [\mu(2\mu + 3\lambda)/(\mu + \lambda)]\varepsilon_{zz}.$$

Es ist üblich, als Proportionalitätsfaktor zwischen Dehnung und Spannung bei einachsiger Beanspruchung den direkt meßbaren *Elastizitätsmodul* E einzuführen, so daß das Hookesche Gesetz in diesem Spezialfall $\sigma_{zz} = E\varepsilon_{zz}$ mit $E = \mu(2\mu + 3\lambda)(\mu + \lambda)$ lautet.

Kompression bedeutet allseitige Zusammendrückung oder allseitige Dehnung unter hydrostatischem Druck p bzw. Zug, d. h., alle Normalspannungskomponenten sind gleich groß, $\sigma_{xx} = \sigma_{yy} = \sigma_{zz} = p$. Der Körper erfährt eine reine Volumenänderung ΔV ohne Gestaltänderung (Abb. 2.2.1–5). Die relative Volumenänderung $\Delta V/V$ ist gleich der Summe

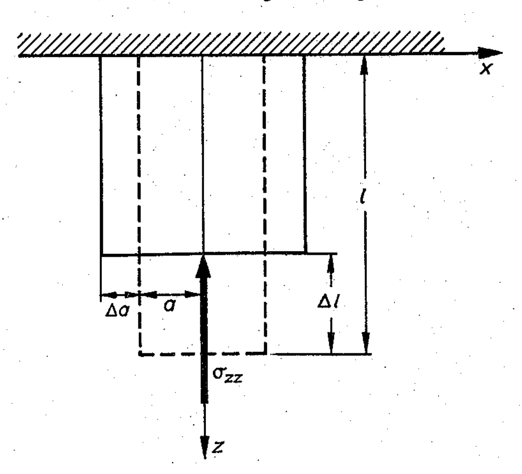


Abb. 2.2.1–4 Einachsiger Druck und Querkontraktion

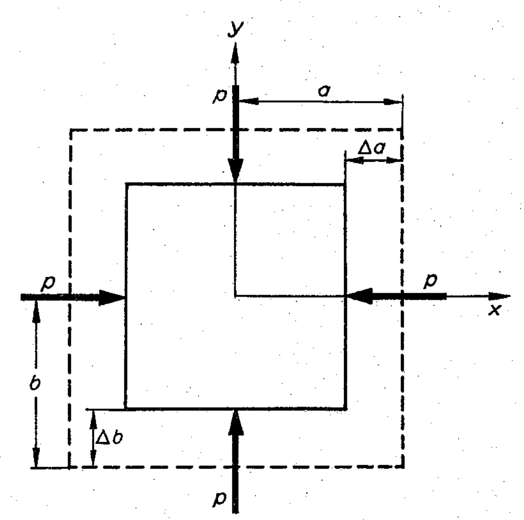


Abb. 2.2.1–5 Kompression durch hydrostatischen Druck

der relativen Längenänderungen, $\Delta V/V = (\Delta a/a) + (\Delta b/b) + (\Delta c/c) = \varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}$. Das letzte Glied entspricht hier der Längenänderung in z-Richtung. Die Längenänderung in einer Richtung bei Anwendung einer Druckspannung setzt sich aus der Verkürzung $-p/E$ infolge des Druckes p in dieser Richtung (vgl. Hookesches Gesetz) und einer Verlängerung $2\nu(p/E)$ zusammen, die von der Querkontraktion herrührt, die der Druck p in den beiden anderen Richtungen verursacht, d.h., es gilt z.B. für die x-Richtung $\Delta a/a = -(p/E) + 2\nu(p/E)$. Die Summe bringt zum Ausdruck, daß sich die Verformungen in den drei Raumrichtungen ungestört überlagern, d. h., es wurde das Superpositionsprinzip (vgl. 2.2.1.1.) angewendet. Werden für die Verformung in y-Richtung $\Delta b/b$ und in z-Richtung $\Delta c/c$ analoge Ausdrücke wie für $\Delta a/a$ angegeben, so ergibt sich $\Delta V/V = -3(1-2\nu)p/E$. Da mit einem Druck $p > 0$ eine Volumenverkleinerung $\Delta V < 0$ verbunden sein muß, ist $1 - 2\nu > 0$ zu fordern. Damit ist aber der Wertebereich der Querkontraktionszahl festgelegt, $0 \leqq \nu \leqq 1/2$. Die obere Grenze bedeutet einen völlig inkompressiblen Körper, d. h., ein Druck ruft keine Volumenänderung hervor. Der Zusammenhang zwischen Druck und Volumenänderung wird kurz durch den *Kompressionsmodul* K ausgedrückt, $p = -K(\Delta V/V)$, dabei ist $K = E/3(1-2\nu) = (2\mu + 3\lambda)/3$.

Biegung gehört ebenfalls zu den Zug- bzw. Druckbeanspruchungen. Bei der Balkenbiegung z. B. wechselt die Beanspruchung von reiner Druckspannung zu reiner Zugspannung. Dazwischen gibt es einen Bereich, in dem die Spannung Null ist, die *neutrale Faser*, die den Schwerpunkt des Balkenquerschnittes enthält. Für Festigkeitsbetrachtungen interessant ist die maximale Zugspannung σ_{max}, die in der äußersten auf Zug belasteten Faser auftritt, $\sigma_{max} = M/W$. Dabei ist M das durch die äußere Belastung des Balkens hervorgerufene Biegemoment, W das *Widerstandsmoment* des Balkens, das nur von Form und Größe des Balkenquerschnittes abhängt.

Das *Flächenträgheitsmoment* I eines Querschnittes bezüglich einer Achse bA ist $I = \iint r^2 \mathrm{d}F$, dabei ist r der Abstand eines Flächenelementes $\mathrm{d}F$ von der Biegeachse bA (Abb. 2.2.1-6). Um sämtliche Flächenelemente des betrachteten Querschnittes zu erfassen, ist über letzteren zu integrieren. Das Widerstandsmoment ist $W = I/e$ mit e als dem Abstand der äußersten Faser des Querschnittes von der Achse bA. In Abb. 2.2.1-7 sind die Flächenträgheitsmomente und die Widerstandsmomente bezüglich der mit bA bezeichneten Achse für einige praktisch wichtige Querschnitte angegeben, wobei in diesen Fällen alle Querschnitte symmetrisch zu bA sind. Ist das nicht der Fall, so ist bei der Berechnung von W aus I zu beachten, daß für die untere und für die obere äußerste Faser verschiedene e einzusetzen sind. Es ist bemerkenswert, daß die Längs- und Querdimensionen des Querschnittes (bezogen auf die Biegeachse bA) mit unterschiedlichem Gewicht in die Berechnung von W eingehen. Da die maximalen Spannungen σ_{max}

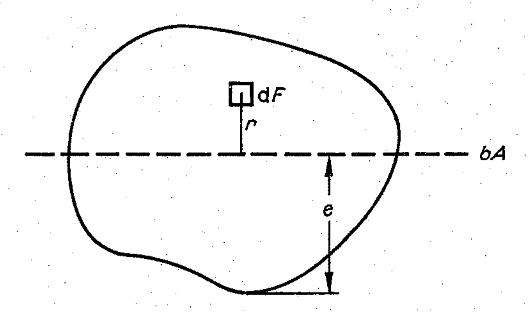


Abb. 2.2.1-6 Berechnung des Flächenträgheitsmomentes

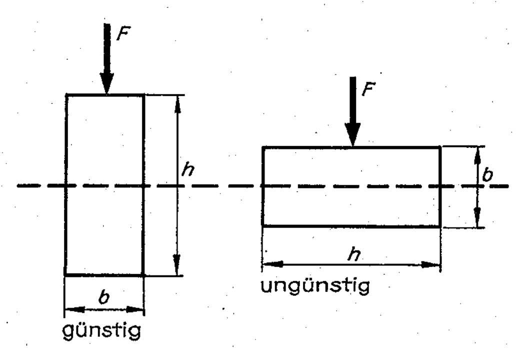


Abb. 2.2.1-8 Mögliche Belastung bei Biegung eines rechteckigen Balkens

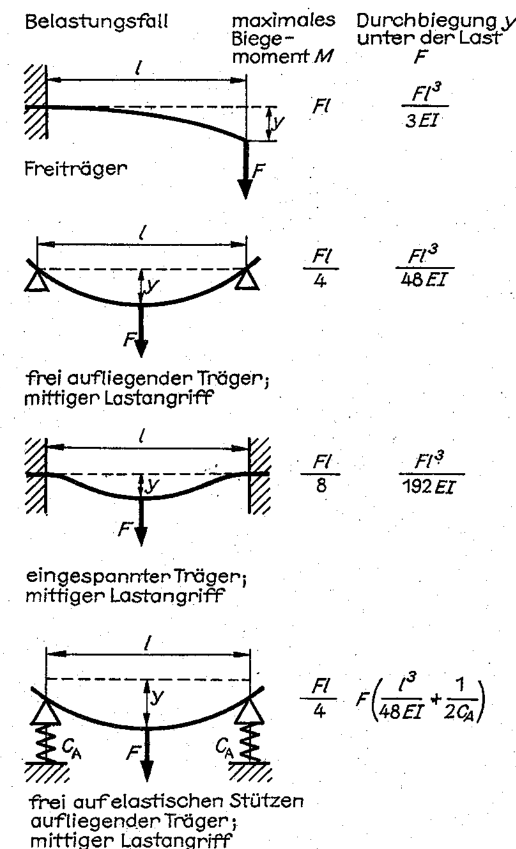


Abb. 2.2.1-9 Biegemomente M und Durchbiegungen y verschieden beanspruchter Balken

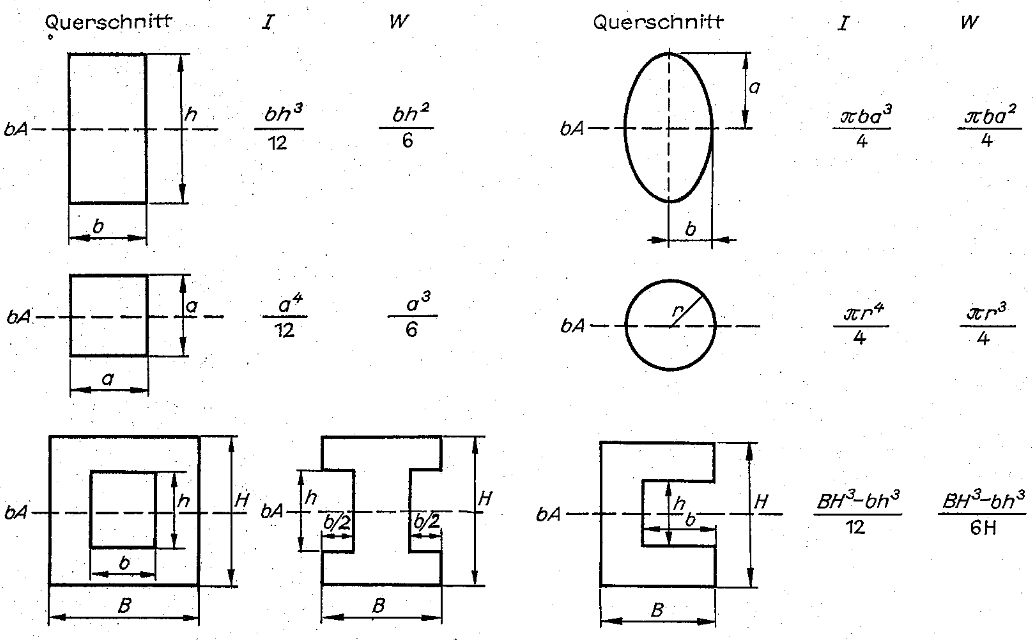


Abb. 2.2.1–7 Flächenträgheitsmomente I und Widerstandsmomente W einiger zur Achse bA symmetrischer Balkenquerschnitte

im gebogenen Balken kleiner sind, wenn das Widerstandsmoment vergrößert wird, sollte die Biegeachse bA bei einem gegebenem Querschnittsprofil immer senkrecht zur größeren Querschnittsabmessung liegen (das gilt allgemein, nicht nur für die speziellen Profile in Abb. 2.2.1–7). Wird z. B. ein Balken mit rechteckigem Querschnitt verwendet und sei $k = h/b$ das Verhältnis von Höhe zu Breite des Balkens ($k > 1$), so ist das Widerstandsmoment in dem Fall, daß die die Biegung verursachende Kraft F parallel zur längeren Seite des Querschnittes angreift (Abb. 2.2.1–8 *links*) $W_l = (1/6)k^{1/2}A^{3/2}$ (A = Querschnittsfläche des Balkens). Wird derselbe Balken parallel zur kürzeren Seite belastet (Abb. 2.2.1–8 *rechts*), so ist $W_k = (1/6)k^{-1/2}A^{3/2}$, d. h., das Verhältnis der beiden Widerstandsmomente ist $W_l/W_k = k$. Es läßt sich somit bei konstanter Querschnittsfläche eines Balkens, und das bedeutet in der Praxis bei unverändertem Materialeinsatz, die Belastbarkeit eines auf Biegung beanspruchten Trägers erhöhen, indem die Form des Querschnittes geeignet gewählt (z. B. Rechteck statt Quadrat) und der Träger so gelagert wird, daß die Spannung σ_{max} den kleinstmöglichen Wert annimmt (vgl. Abb. 2.2.1–8).

Außer W muß zur Berechnung der maximalen Biegespannung des Balkens σ_{max} das von der äußeren Last F erzeugte *maximale Biegemoment M* bekannt sein. Dieses ist außer von der Größe F und der Balkenlänge l von den Auflagebedingungen des Balkens abhängig. In Abb. 2.2.1–9 sind die Biegemomente für einige Lagerungsfälle angegeben. Zusätzlich ist der *Biegepfeil* aufgeführt, d. h. die Durchbiegung y des Balkens am Lastangriffspunkt, die bei den hier mitgeteilten einfachen Fällen gleichzeitig die größte Durchbiegung ist. Da es sich bei der Biegung um einen speziellen, wenn auch relativ komplizierten Fall der Zug- bzw. Druckbeanspruchung handelt, tritt als Materialkonstante, die die elastischen Eigenschaften charakterisiert, der Elastizitätsmodul E auf (I = Flächenträgheitsmoment) (vgl. Abb. 2.2.1–7). Im letzten Lagerungsfall der Abb. 2.2.1–9 wird angenommen, daß beide Auflager die gleiche Federkonstante c_A haben.

Scherung bedeutet eine reine Gestaltänderung des Körpers ohne eine Volumenänderung. Greift an einem rechteckigen Quader eine Schubspannung τ_{xy} an (Abb. 2.2.1–10), so werden die einzelnen Schichten des Quaders so gegeneinander verschoben, daß die vorher senkrechten Begrenzungsflächen nun mit der Vertikalen einen Winkel γ_{xy} bilden, der bei kleinen Verformungen wegen $\tan \gamma_{xy} = \Delta y/h \approx \gamma_{xy}$ gleich der Schiebung ist. Der Zusammenhang zwischen Schubspannung und Schiebung folgt aus dem allgemein formulierten Hookeschen Gesetz, $\tau_{xy} = \mu\gamma_{xy}$. Es ist üblich, das Verhältnis τ_{xy}/γ_{xy} bei reiner Scherung als *Schub-* oder *Schermodul G* zu bezeichnen, d. h., $G = \mu$.

Wird ein Draht oder Stab durch ein äußeres Drehmoment verdrillt (tordiert), so verschieben sich die einzelnen Querschnitte gegeneinander. Denkt man sich auf dem Stabum-

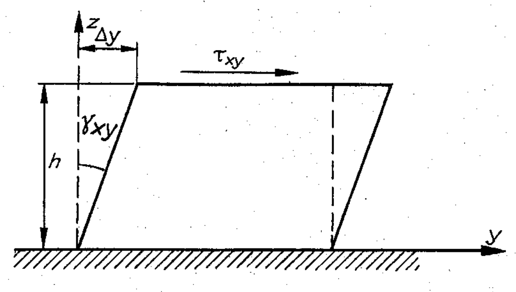


Abb. 2.2.1–10 Scherung und Schubspannung

fang zur Stabachse parallele Geraden markiert, so werden diese infolge der *Torsion* gegen die Achse um den Winkel γ geneigt, der mit der Schiebung identisch ist. Der Verdrehwinkel α eines aus dem Stab herausgegriffenen Hohlzylinders mit dem Radius r bis $\alpha = l\gamma/r$ (l = Länge des Zylinders). Unter Verwendung des Hookeschen Gesetzes für γ wird das Drehmoment M, das zur *Verdrillung* eines Vollstabes des Radius R und der Länge l um den Winkel α erforderlich ist, $M = (\pi/2)\,(GR^4/l)\alpha$ (G = Schubmodul des Stabmateriales). Wegen dieser engen Verbindung von Torsion und Scherung wird G auch als *Torsionsmodul* bezeichnet, insbesondere, da die Verdrillung des Drahtes in der Praxis häufig dazu benutzt wird, G experimentell zu bestimmen. Andererseits wird in Meßinstrumenten aus der auf einer Skala angezeigten Größe der Verdrillung eines Drahtes das Torsionsmoment M bestimmt und damit der Wert der solch ein Moment hervorrufenden Größe gemessen, z. B. Messung des elektrischen Stromes durch Galvanometer.
Wird eine eng gewickelte *Schraubenfeder* auf Zug beansprucht, so bedeutet das, daß der Draht tordiert wird, d. h., die Elastizität des Federmateriales wird durch den Schubmodul G und nicht wie bei der Dehnung durch den Elastizitätsmodul charakterisiert. Die Verlängerung w der Feder infolge einer axial wirkenden Kraft F ist $w = (4a^3n/GR^4)F$ (a = Wickelradius der Feder, R = Radius des Drahtes, n = Zahl der Windungen). Damit ergibt sich die *Federkonstante* c der Schraubenfeder $c = F/w = GR^4/4a^3n$, d. h., sie läßt sich aus den konstruktiven Daten der Feder berechnen. Zur Dimensionierung einer Feder ist noch die maximale Schubspannung τ_{max} im Drahtquerschnitt bei an der Feder wirkender Kraft F von Interesse, $\tau_{max} = [2a/(\pi R^3)]F$.
Zusammenhang zwischen Elastizitätszahlen. Da die elastischen Eigenschaften eines isotropen Mediums durch zwei Elastizitätszahlen hinreichend charakterisiert sind, müssen zwischen den einzelnen Moduln, die zur Kennzeichnung der Abhängigkeit der Spannung von der Verformung bei einfachen und übersichtlichen Belastungsbedingungen eingeführt wurden, Abhängigkeiten bestehen. Die Kenntnis von zwei Moduln reicht also aus, alle anderen zu berechnen. Der Zusammenhang von ν, E, K und G mit den Laméschen Parametern μ und λ wurde bereits angegeben. Daraus folgen $K = E/[3(1-2\nu)]$ und $G = E/[2(1+\nu)]$.

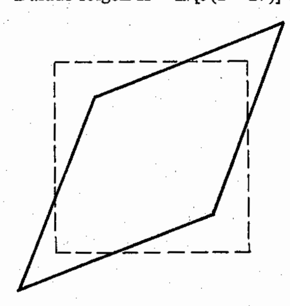


Abb. 2.2.1–11 Scherung als Überlagerung von Dehnung und Querkontraktion

Diese Beziehung $E = 2G(1+\nu)$ läßt sich qualitativ auch so verstehen, daß eine Scherung aufgefaßt werden kann als die Überlagerung einer Dehnung und einer Querkontraktion. Der Quader in Abb. 2.2.1–11 erfährt dadurch seine Gestaltänderung, daß die Diagonale gedehnt wird, was gleichzeitig mit einer Querkontraktion verbunden ist, d. h., G muß aus E und ν berechenbar sein. Wegen $0 \leq \nu \leq 1/2$ folgt als Wertebereich für G aus obiger Beziehung $E/3 \leq G \leq E/2$. Die beiden Gleichungen für E und K, in denen alle vier Konstanten E, K, G und ν enthalten sind, ermöglichen so die Berechnung aller Elastizitätszahlen, wenn zwei experimentell bekannt sind. In der Tab. 2.2.1–12 sind für einige ausgewählte Materialien die entsprechenden Werte angegeben, die auch eine Überprüfung der genannten Zusammenhänge erlauben. Es kann sich dabei natürlich nur um Orientierungswerte handeln, da die Elastizitätszahlen auch durch die Zusammensetzung des Werkstoffes beeinflußt werden, so daß bei exakten Rechnungen auf einschlägige Tabellenwerke zurückzugreifen ist.

Tab. 2.2.1–12 Elastizitätszahlen einiger wichtiger Werkstoffe

	Elastizitätsmodul E in $10^{10}\,Nm^{-2}$	Torsionsmodul G in $10^{10}\,Nm^{-2}$	Querkontraktionszahl ν	Kompressionsmodul K in $10^{10}\,Nm^{-2}$
Stahl	20...22	7,5...9	0,2...0,3	11...20
Gußeisen	7,5	3	0,25	5
Messing	7...11	2,5...4	0,3...0,4	5...19
Aluminium	7,3	2,7	0,34	7,6
Kupfer	12,5	4,6	0,35	13,9
Glas	5...9	2...4	0,2...0,3	3...8
Holz in Faserrichtung	0,8...1,6	–	0,1...0,2	–

2.2.1.3. Anisotrope Medien. Der allgemeine lineare Zusammenhang (verallgemeinertes Hookesches Gesetz) ergab (vgl. 2.2.1.1.) für die Abhängigkeit der sechs Spannungskomponenten von den sechs Verformungskomponenten $6 \times 6 = 36$ Konstanten c_{ik} ($i, k = 1, 2, 3, 4, 5, 6$). Für einen Festkörper, der linearelastisch verformt wird, ist die elastische Energie eine Zustandsgröße (vgl. 2.6.), d. h., die Arbeit, die erforderlich ist, den Körper zu deformieren, hängt nur vom Anfangs- und vom Endzustand des Körpers ab, nicht aber von dem Weg, auf dem die Deformation erfolgt. Die mathematische Bedingung für die Existenz der Zustandsgröße fordert $c_{ik} = c_{ki}$, d. h., die 30 Konstanten mit gemischten Indizes ($i \neq k$) sind paarweise gleich, so daß sich auch im allgemeinsten Fall die Zahl der unabhängigen elastischen Konstanten auf $36 - 30/2 = 21$ reduziert. Alle Kristalle verhalten sich mechanisch anisotrop. Das hängt mit der Kristallstruktur (vgl. 6.1.) zusammen, da durch die regelmäßige Anordnung der Gitterbauteilchen die Zahl der Gitterpunkte je Längeneinheit in verschiedenen Raumrichtungen verschieden voneinander ist. Diese prinzipielle Anisotropie wirkt sich bei makroskopischen Verformungsexperimenten allerdings nur aus, wenn die Messung an Einkristallen (vgl. 6.2.) vorgenommen wird, da sich bei polykristallinem Gefüge wegen der regellosen Anordnung der einzelnen Kristallite die Richtungsabhängigkeit herausmittelt, so daß diese Materialien wie mechanisch isotrope Medien (vgl. 2.2.1.2.) behandelt werden können.
Die *Kristallsymmetrie* (vgl. 6.1.) bewirkt, daß nicht alle Richtungen im Kristall verschieden voneinander sind. Das bedeutet, daß die Zahl der unabhängigen elastischen Konstanten c_{ik} weiter verringert wird, wobei die endgültige Anzahl von den Symmetrieelementen des betreffenden Kristallsystemes abhängt, wie an einigen einfachen Beispielen demonstriert werden soll. Um die Transformation, die mit den entsprechenden Symmetrieoperationen verbunden ist, durchführen zu können, wird nach Voigt für die Indizes i

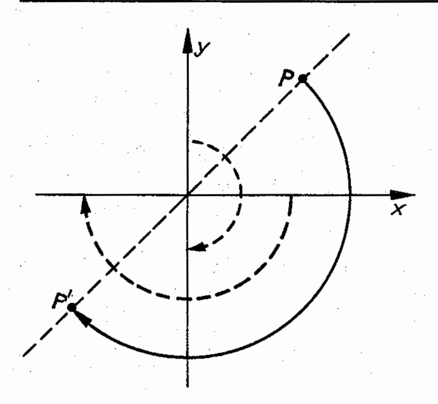


Abb. 2.2.1–13 Eine zweizählige Drehachse führt P in den identischen Punkt P' über, das entspricht der Transformation $x \to -x;\ y \to -y;\ z \to z$

und k der c_{ik} folgende Zuordnung zu den Koordinaten x, y und z getroffen:

$1 \equiv xx;\ 2 \equiv yy;\ 3 \equiv zz;\ 4 \equiv yz = zy;\ 5 \equiv zx = xz;\ 6 \equiv xy = yx.$

Die einzelnen Konstanten c_{ik} transformieren sich so, wie sich das entsprechend der angegebenen Vorschrift zugehörige vierfache Produkt der Koordinaten transformiert, z. B. wird c_{14} transformiert wie das Koordinatenprodukt $xxyz$. Der Übersichtlichkeit halber werden die 21 unabhängigen elastischen Konstanten in folgendem Schema angeordnet:

$$\begin{array}{cccccc} c_{11} & c_{12} & c_{13} & c_{14} & c_{15} & c_{16} \\ & c_{22} & c_{23} & c_{24} & c_{25} & c_{26} \\ & & c_{33} & c_{34} & c_{35} & c_{36} \\ & & & c_{44} & c_{45} & c_{46} \\ & & & & c_{55} & c_{56} \\ & & & & & c_{66} \end{array}$$

Diese Anordnung widerspiegelt gleichzeitig das Koeffizientenschema des verallgemeinerten Hookeschen Gesetzes, wenn man sie sich unterhalb der Diagonale symmetrisch zu dieser vervollständigt denkt. Eine erlaubte Symmetrieoperation des Kristalles bedeutet, daß sich infolge dieser Operation seine Eigenschaften nicht ändern, d. h., elastische Konstanten im Koeffizientenschema nach der Symmetrieoperation müssen gleich denen sein, die im Schema an der gleichen Stelle vor der Operation standen. Das *trikline Kristallsystem* enthält Klassen mit und ohne *Inversionszentrum* (vgl. 6.1.). Da sich nach dem Prinzip von Voigt aber die Symmetrie des Vorganges der Symmetrie des Kristalles überlagert, die elastische Deformation aber immer ein Inversionszentrum hat (die Dehnung in einer bestimmten Richtung ist gleich der in der entgegengesetzten), unterscheiden sich diese Klassen nicht hinsichtlich der Zahl der elastischen Konstanten. Eine *Inversion* (Punktspiegelung) bedeutet, daß sich die Vorzeichen aller Koordinaten umkehren, d. h., es findet folgende Koordinatentransformation statt: $x \to -x;\ y \to -y;\ z \to -z$. Das bedeutet aber, daß sich alle Indizes $i, k = 1, 2, 3, 4, 5, 6$ identisch transformieren. Dasselbe gilt deshalb auch für die elastischen Konstanten, so daß das Koeffizientenschema nach einer Inversion unverändert bleibt, d. h., im triklinen Kristallsystem treten 21 unabhängige Konstanten auf. Das *monokline Kristallsystem* hat eine *zweizählige Drehachse*, die als z-Achse gewählt wird, d. h., die beiden Punkte P und P' in Abb. 2.2.1–13, die nach einer Drehung um 180° ineinander übergehen, sind identische Gitterpunkte. Dasselbe gilt für alle anderen Gitterpunkte. Diese Drehung ist aber identisch mit der Koordinatentransformation $x \to -x;\ y \to -y;\ z \to z$, wie sofort aus Abb. 2.2.1–13 abzulesen ist. Die sich daraus ergebende Transformation der Indizes ist

$1 \equiv xx \to (-x)(-x) = xx \equiv 1;\quad 2 \to 2;\quad 3 \to 3;\quad 4 \equiv yz \to (-y)z = -yz \equiv -4;\ 5 \to -5;\ 6 \to 6.$ Für die elastische Konstante c_{14} z. B. bedeutet das die Transformation in $-c_{14}$, da 14 übergeht in $1(-4) = -14$. Werden nach dieser Vorschrift alle Konstanten transformiert, so lautet das Koeffizientenschema nach der Symmetrieoperation

$$\begin{array}{cccccc} c_{11} & c_{12} & c_{13} & -c_{14} & -c_{15} & c_{16} \\ & c_{22} & c_{23} & -c_{24} & -c_{25} & c_{26} \\ & & c_{33} & -c_{34} & -c_{35} & c_{36} \\ & & & c_{44} & c_{45} & -c_{46} \\ & & & & c_{55} & -c_{56} \\ & & & & & c_{66} \end{array}$$

Da dieses aber gleich dem vor der Symmetrieoperation sein muß, ergeben sich außer den identisch erfüllten Beziehungen die Gleichungen $c_{14} = -c_{14}$, $c_{15} = -c_{15}$, $c_{24} = -c_{24}$, $c_{25} = -c_{25}$, $c_{34} = -c_{34}$, $c_{35} = -c_{35}$, $c_{46} = -c_{46}$, $c_{56} = -c_{56}$, die nur dann zu erfüllen sind, wenn alle in diesen Beziehungen auftretenden Konstanten Null sind, so daß das Koeffizientenschema für das monokline Kristallsystem

$$\begin{array}{cccccc} c_{11} & c_{12} & c_{13} & 0 & 0 & c_{16} \\ & c_{22} & c_{23} & 0 & 0 & c_{26} \\ & & c_{33} & 0 & 0 & c_{36} \\ & & & c_{44} & c_{45} & 0 \\ & & & & c_{55} & 0 \\ & & & & & c_{66} \end{array}$$

lautet, d. h., im monoklinen System existieren nur noch 13 unabhängige elastische Konstanten. Eine zweite zweizählige Drehachse senkrecht zu der vorigen (Koordinatentransformation $x \to -x;\ y \to y;\ z \to -z$, wenn die y-Achse diese zweite Drehachse ist), wie sie im *rhombischen Kristallsystem* auftritt, vermindert nach einem entsprechenden Schema die Zahl der unabhängigen Konstanten auf neun, das Koeffizientenschema lautet jetzt

$$\begin{array}{cccccc} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ & c_{22} & c_{23} & 0 & 0 & 0 \\ & & c_{33} & 0 & 0 & 0 \\ & & & c_{44} & 0 & 0 \\ & & & & c_{55} & 0 \\ & & & & & c_{66} \end{array}$$

Auf diese Art und Weise kann unter Beachtung der erlaubten Symmetrieoperationen für jedes Kristallsystem das Koeffizientenschema ermittelt werden. Das *kubische Kristallsystem* ist durch die Existenz von drei senkrecht aufeinanderstehenden *vierzähligen Drehachsen* charakterisiert, was natürlich die Symmetrien des rhombischen Systems mit einschließt, so daß bei der Behandlung des kubischen Systems vom Koeffizientenschema des letzteren ausgegangen werden kann. Wird als eine der vierzähligen Drehachsen die z-Achse gewählt, so entspricht der Überführung des Punktes P in den identischen Punkt P' in Abb. 2.2.1–14 nach einer Drehung um 90° die Koordinatentransformation $x \to -y;\ y \to x;\ z \to z$. Damit folgt für die Indizes $1 \equiv xx \to (-y)(-y) = yy \equiv 2;\quad 2 \to 1;\quad 3 \to 3;\quad 4 \to 5;$

$5 \equiv zz \to z(-y) = -zy \equiv -4$; $6 \to -6$; d.h., es transformiert sich z. B. c_{13} in c_{23} usw., so daß das Koeffizientenschema nach dieser Drehung

$$\begin{matrix} c_{22} & c_{12} & c_{23} & 0 & 0 & 0 \\ & c_{11} & c_{13} & 0 & 0 & 0 \\ & & c_{33} & 0 & 0 & 0 \\ & & & c_{55} & 0 & 0 \\ & & & & c_{44} & 0 \\ & & & & & c_{66} \end{matrix}$$

lautet. Gleichsetzen mit dem vorhergehenden Schema liefert $c_{11} = c_{22}$; $c_{13} = c_{23}$; $c_{44} = c_{55}$. Wird das sich so ergebende Koeffizientenschema

$$\begin{matrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ & c_{11} & c_{13} & 0 & 0 & 0 \\ & & c_{33} & 0 & 0 & 0 \\ & & & c_{44} & 0 & 0 \\ & & & & c_{44} & 0 \\ & & & & & c_{66} \end{matrix}$$

noch entsprechend einer Drehung um 90° um die x-Achse transformiert ($x \to x$; $y \to z$; $z \to -y$) und mit dem sich ergebenden Schema gleichgesetzt, so ergibt sich endlich als Schema der elastischen Konstanten für das kubische Kristallsystem

$$\begin{matrix} c_{11} & c_{12} & c_{12} & 0 & 0 & 0 \\ & c_{11} & c_{12} & 0 & 0 & 0 \\ & & c_{11} & 0 & 0 & 0 \\ & & & c_{44} & 0 & 0 \\ & & & & c_{44} & 0 \\ & & & & & c_{44} \end{matrix}$$

Eine weitere Verminderung der elastischen Konstanten ist auch bei Berücksichtigung der dritten vierzähligen Drehachse (y-Achse) nicht möglich, d. h., bei einem kubischen Kristall existieren drei unabhängige elastische Konstanten; er verhält sich also mechanisch nicht wie ein elastisch isotropes Medium. Mit diesem Koeffizientenschema läßt sich das Hookesche Gesetz für einen Kristall des kubischen Systems angeben:

$$\sigma_{xx} = c_{11}\varepsilon_{xx} + c_{12}\varepsilon_{yy} + c_{12}\varepsilon_{zz},$$
$$\sigma_{yy} = c_{12}\varepsilon_{xx} + c_{11}\varepsilon_{yy} + c_{12}\varepsilon_{zz},$$
$$\sigma_{zz} = c_{12}\varepsilon_{xx} + c_{12}\varepsilon_{yy} + c_{11}\varepsilon_{zz},$$
$$\tau_{yz} = c_{44}\gamma_{yz}; \quad \tau_{zx} = c_{44}\gamma_{zx}; \quad \tau_{xy} = c_{44}\gamma_{xy}.$$

Der Vergleich mit dem Hookeschen Gesetz eines isotropen Mediums (vgl. 2.2.1.2.) fordert für letzteres $c_{11} = 2\mu + \lambda$; $c_{12} = \lambda$ und $c_{44} = \mu$ (λ, μ =Lamésche Konstanten). Damit sich ein kubischer Kristall isotrop verhält, müßte deshalb zwischen seinen elastischen Konstanten die Beziehung $c_{44} = (c_{11} - c_{12})/2$ gelten. Wie die Beispiele in Tab.2.2.1–15 für einige kubische Einkristalle zeigen, ist das nur in wenigen Ausnahmefällen angenähert erfüllt.

2.2.2. Dynamisches elastisches Verhalten von Festkörpern

Ein Festkörper wird dynamisch beansprucht, wenn sich z.B. ein Impuls durch diesen ausbreitet. Die Fortpflanzung des Impulses wird daher außer von den elastischen Eigenschaften des Materiales wegen der mit der Ausbreitung verbundenen Verzerrungen auch von der Trägheit der bewegten Teilchen beeinflußt. Letzteres kommt dadurch zum Ausdruck, daß im folgenden neben den elastischen Konstanten als Materialparameter die Dichte ϱ des Festkörpers auftreten wird. Die weitere Behandlung wird auf isotrope Medien beschränkt, so daß die dafür charakteristischen Moduln E, G und ν benutzt werden (vgl. 2.2.1.2.).

Kann der Festkörper als unbegrenzt angesehen werden, so breitet sich eine beliebige Störung als *fortschreitende Welle* mit der für den Wellentyp charakteristischen Geschwindigkeit c aus (vgl. 2.3.3.). Die Frequenz ν der Welle ist gleich der der Erregung und die Wellenlänge λ folgt aus $\lambda = c/\nu$ (vgl. 2.3.3.). Eine gewisse Ausnahme bilden hier lediglich die Biegungswellen (vgl. 2.2.2.1.). Ist das Medium begrenzt, so wird die durch eine Störung ausgelöste ankommende Welle reflektiert. Hin- und rücklaufende Welle interferieren und bilden eine *stehende Welle* (vgl. 2.3.6.), deren Phase von den Randbedingungen an der Begrenzung abhängt.

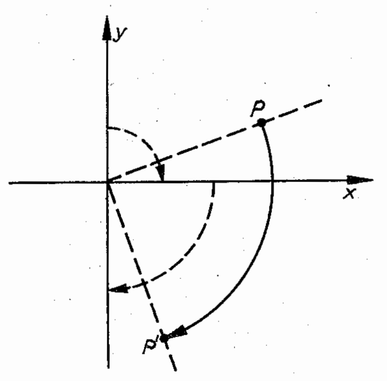


Abb. 2.2.1–14 Eine vierzählige Drehachse führt P in den identischen Punkt P' über, das entspricht der Transformation $x \to -y$; $y \to x$; $z \to z$

Tab. 2.2.1–15 Elastische Konstanten einiger ausgewählter kubischer Einkristalle (jeweils in $10^{10}\,\mathrm{Nm^{-2}}$)

	c_{11}	c_{12}	c_{44}	$(c_{11} - c_{12})/2$
Eisen	23,3	13,5	11,8	4,9
Aluminium	10,7	6,1	2,8	2,3
Kupfer	16,8	12,1	7,5	2,4
Wolfram	52,3	20,5	16,1	15,9
Silicium	16,6	6,4	8,0	5,1
Diamant	107,6	12,5	57,6	47,6
Steinsalz	4,9	1,25	1,27	1,83

Im Gegensatz zu den Flüssigkeiten und Gasen, in denen sich nur *Longitudinalwellen* (vgl. 2.3.2.) ausbreiten können, treten in Festkörpern auch *Transversalwellen* (vgl. 2.3.2.) auf. Das ist möglich, da letztere neben der reinen Volumenelastizität, die in allen Medien auftritt, auch die elastische Verformungsmöglichkeit der Scherung (vgl. 2.2.1.2.) aufweisen, bei der nur die Gestalt geändert wird.

2.2.2.1. Eindimensionale Wellenausbreitung. Die Auslenkung der Teilchen des Festkörpers wird in diesem Fall sowohl für die longitudinalen als auch für die transversalen Wellen durch die einfache *eindimensionale Wellengleichung* (vgl. 2.3.3.) beschrieben. Eine Ausnahme bilden die Biegungswellen (s. u.). Eine *Longitudinalwelle* kann sich sowohl

in einem Stab als auch in einer gespannten Saite ausbreiten, in letzterem Fall durch Reibung in Längsrichtung mittels eines mit Kolophonium eingeriebenen Lappens. In beiden Fällen ist die *Wellengeschwindigkeit* $c_{l,1} = (E/\varrho)^{1/2}$ nur von der Dichte ϱ und vom Elastizitätsmodul E abhängig, d. h., die Ausbreitungsgeschwindigkeit einer Longitudinalwelle in einer Saite ist unabhängig von Spannkraft und Querschnitt der Saite. Als charakteristischer elastischer Parameter erscheint hier E, da die Ausbreitung einer Longitudinalwelle in einer Folge von Dehnungen und Stauchungen besteht. In Tab.2.2.2–1 sind für einige Materialien die Ausbreitungsgeschwindigkeiten $c_{l,1}$ angegeben, mit denen sich auch der Schall (vgl. 2.3.8.) in Stäben ausbreitet. Diese Geschwindigkeiten sind etwa eine Größenordnung größer als die *Schallgeschwindigkeit* in Luft. *Transversalwellen* in Stäben breiten sich als *Torsionswellen* aus, ihre Ausbreitungsgeschwindigkeit $c_t = (G/\varrho)^{1/2}$ wird durch den Torsionsmodul G bestimmt. Da G immer kleiner als E ist (vgl. 2.2.1.2.), gilt auch $c_t < c_{l,1}$. Einige Beispiele für c_t sind in Tab. 2.2.2–1 angegeben. Die Ausbreitungsgeschwindigkeit der Transversalwelle einer gespannten Saite c_s ist durch die Spannung $\sigma_s = F_s/A_s$ bestimmt, $c_s = (\sigma_s/\varrho)^{1/2}$ (F_s = Spannkraft, A_s = Querschnitt der Saite). In allen genannten Fällen tritt keine *Dispersion* auf, d.h., die Ausbreitungsgeschwindigkeit der Wellen ist unabhängig von der Wellenlänge.

Sind die Festkörper beidseitig begrenzt, so bilden sich nach einer Störung stehende Wellen als Eigenschwingungen aus. Es sind verschiedene Schwingungsformen möglich. Die stehende Welle mit der größten Wellenlänge λ_0 wird als *Grundschwingung* bezeichnet. Es folgen mit abnehmenden Wellenlängen $\lambda_1 > \lambda_2 > \lambda_3 \ldots$ die erste, zweite, dritte *Oberschwingung* usw. Da sich an einem festen Ende immer ein Schwingungsknoten, an einem freien Ende aber immer ein Schwingungsbauch bei der Reflexion der Welle ergibt, lassen sich die für eine Eigenschwingung möglichen Wellenlängen sofort angeben, wenn diese Bedingungen beachtet werden. In Abb. 2.2.2–2 sind für einige wichtige einfache Fälle die Wellenlängen der Grundschwingung sowie der ersten und zweiten Oberschwingung abhängig von der Länge l zwischen den beiden Enden angegeben. Für die k-te Eigenschwingung gilt ($k = 0, 1, 2 \ldots$): λ_k(fest-fest) $= 2l/(k+1)$; λ_k(fest-frei) $= 4l/(2k+1)$; λ_k(frei-frei) $= 2l/(k+1)$.

Die entsprechenden *Eigenfrequenzen* ν_k ergeben sich aus $\nu_k = c/\lambda_k$ (vgl.2.3.3.), wobei für c die Ausbreitungsgeschwindigkeit für den jeweils interessierenden Wellentyp einzusetzen ist, d. h. $c_{l,1}$ bei Longitudinalschwingungen c_t bei Torsionsschwingungen im Stab und c_s bei Transversalschwingungen der gespannten Saite. In allen Fällen sind die Frequenzen der Oberschwingungen ganzzahlige Vielfache der Frequenz der Grundschwingung.

Biegungswellen in Stäben oder Balken, bei denen die Stabelemente transversal schwingen, sind nicht zu verwechseln mit den oben behandelten eigentlichen Transversalwellen im Stab, den Torsionswellen, bei denen das Material auf Scherung beansprucht wird. Da es sich bei der Biegung um eine spezielle Art der Dehnung handelt (vgl. 2.2.1.2.), wird die Ausbreitungsgeschwindigkeit durch den Elastizitätsmodul E des Stabmateriales bestimmt. Bemerkenswert ist, daß die Auslenkung der Stabelemente nicht mehr durch die übliche Wellengleichung (partielle Differentialgleichung zweiter Ordnung, vgl. 2.3.3.), sondern durch eine partielle Differentialgleichung vierter Ordnung beschrieben wird. Das hat

Tab. 2.2.2–1 Ausbreitungsgeschwindigkeiten elastischer Wellen in $10^3\,\mathrm{ms^{-1}}$ in einigen polykristallinen Metallen und in Glas

	Longitudinalwelle eindimensional $c_{l,1}$	Longitudinalwelle dreidimensional $c_{l,3}$	Transversalwelle c_t
Stahl	5,0…5,3	5,3…6,1	3,1…3,4
Messing	2,9…3,6	3,3…5,3	1,7…2,2
Aluminium	5,2	6,5	3,2
Kupfer	3,7	4,7	2,3
Glas	4,5…6,1	4,8…7,1	2,9…4,1

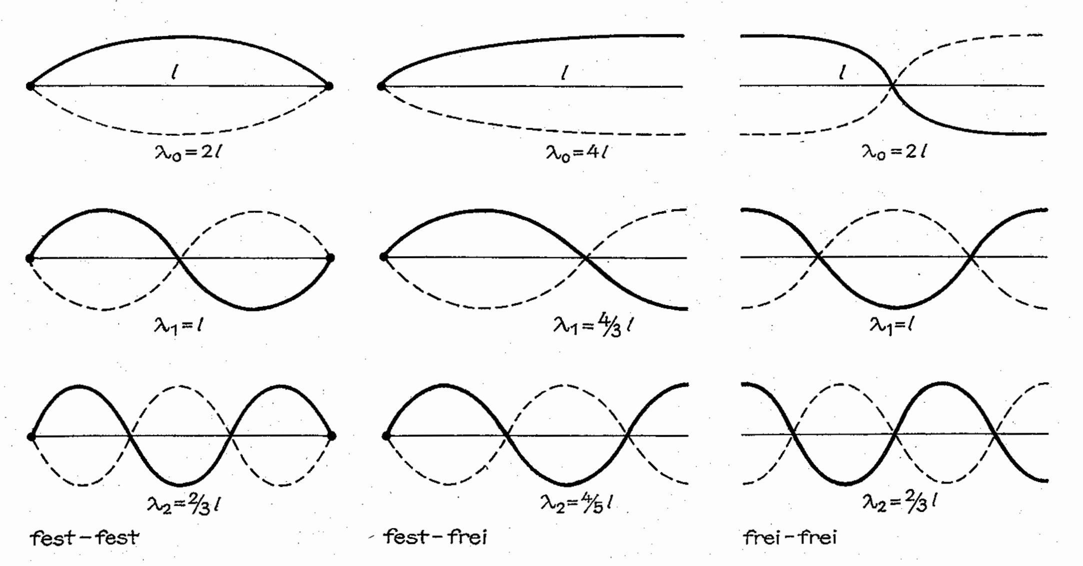
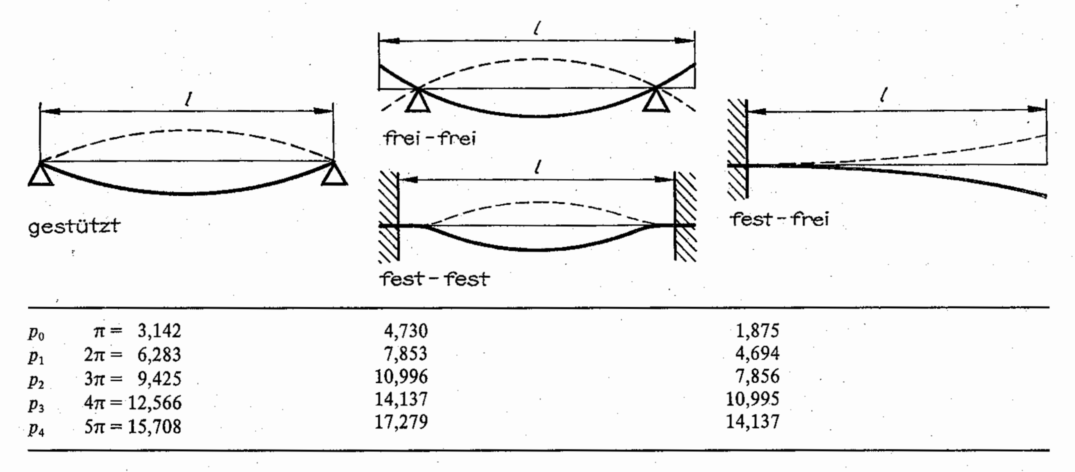


Abb. 2.2.2–2 Grundschwingung sowie erste und zweite Oberschwingung bei verschiedenen Randbedingungen

Tab. 2.2.2–3 Faktor p_k für verschiedengelagerte Balken



	gestützt	frei–frei / fest–fest	fest–frei
p_0	$\pi =$ 3,142	4,730	1,875
p_1	$2\pi =$ 6,283	7,853	4,694
p_2	$3\pi =$ 9,425	10,996	7,856
p_3	$4\pi =$ 12,566	14,137	10,995
p_4	$5\pi =$ 15,708	17,279	14,137

einmal zur Folge, daß sich nicht mehr beliebige, sondern nur noch harmonische Störungen unbeeinflußt ausbreiten können. Zum anderen wird die Ausbreitungsgeschwindigkeit $c_B = (2\pi/\lambda)\,(EI/A\varrho)^{1/2}$ abhängig von der Wellenlänge λ, d. h., es tritt jetzt *Dispersion* auf (I = Flächenträgheitsmoment, A = Querschnittsfläche, ϱ = Dichte des Stabes bzw. Balkens). Zum dritten lassen sich die Wellenlänge von Grund- und Oberschwingungen bei beidseitiger Begrenzung des Stabes nicht mehr so einfach wie in den oben behandelten Fällen angeben, da jetzt an jedem Ende die Erfüllung zweier Randbedingungen gefordert werden muß, was nur noch durch mathematische Lösung des Eigenwertproblems erreicht werden kann. Beispielsweise zeigt der Vergleich eines beidseitig eingespannten Balkens (vgl. Abb. 2.2.1–9) mit einer beidseitig eingespannten Saite (vgl. Abb. 2.2.2–2 *links*), daß in beiden Fällen an den Einspannungen die Auslenkung Null sein muß, daß aber zusätzlich wegen der Biegesteifigkeit des Balkens gefordert werden muß, daß im ersten Fall die Biegelinie in die Einspannungen mit einer horizontalen Tangente einlaufen muß. Die Berechnung liefert als Eigenfrequenz der k-ten Eigenschwingung $\nu_k = (p_k^2/2\pi l^2)\,(EI/A\varrho)^{1/2}$, wobei der Faktor p_k für verschiedengelagerte Balken der Länge l bis zur vierten Oberschwingung in Tab. 2.2.2–3 angegeben ist. Es ist zu beachten, daß bei dem „gestützten" Balken die Balkenenden auf den Stützen frei aufliegen, während bei dem „frei-freien" Balken die Balkenenden frei schwingen und der Ort der Unterstützung so gewählt werden muß, daß sich dort ein Knoten befindet (bei der Grundschwingung z. B. beträgt der Abstand des Knotens vom freien Ende $0{,}224 \cdot l$). Der beiderseits fest eingespannte Balken hat die gleichen Eigenfrequenzen wie der beiderseits freie. Die Frequenzen der Oberschwingungen sind keine ganzzahligen Vielfachen der Frequenz der Grundschwingung mehr, sie nehmen stärker zu als z. B. die Frequenzen der Transversalschwingungen der Saite. Werden Flächenträgheitsmoment und Querschnittsfläche für einen rechteckigen Balken eingesetzt, der in Richtung seiner Querdimension h schwingt (in Abb. 2.2.2–4 ist das für einen einseitig eingespannten Balken skizziert), so wird $\nu_k = (p_k^2 h/4\pi l^2)\,(E/3\varrho)^{1/2}$, d. h., die Eigenfrequenzen sind unabhängig von der zur Schwingungsrichtung senkrechten Querausdehnung b des Balkens.

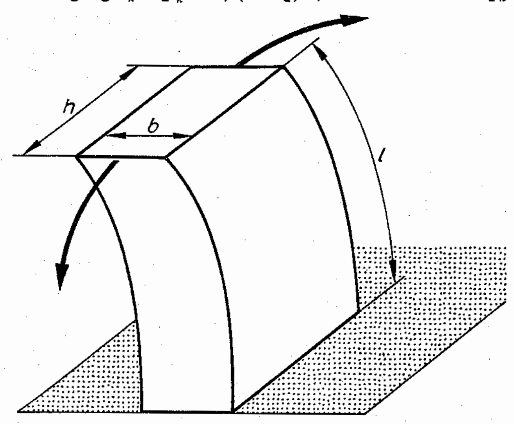


Abb. 2.2.2–4 Rechteckiger Balken parallel zu seiner Querdimension h schwingend

2.2.2.2. Zweidimensionale Wellenausbreitung. Bei den Medien mit zweidimensionaler Wellenausbreitung handelt es sich um *Membranen* und *Platten*. Es können sich sowohl Longitudinal- als auch Transversalwellen ausbilden, aber nur letztere haben eine praktische Bedeutung.

Membranen sind das Analogon zur Saite im eindimensionalen Fall. Da sie keine Biegesteifigkeit besitzen, müssen sie in einem Rahmen mit einer Spannung σ_M eingespannt werden. Deshalb ist bei Membranen der Rand auch stets eine Knotenlinie, d. h., längs des Rahmens können keine Verrückungen auftreten. Für eine rechteckige Membran (Seitenlängen a und b) lassen sich die Eigenfrequenzen ν_{hk} explizit angeben:

$$\nu_{hk} = (1/2)\{(1+h)^2/a^2 + (1+k)^2/b^2\}^{1/2}\,(\sigma_M/\varrho)^{1/2}.$$

Hier ist ϱ die Dichte des Membranmateriales, h und k sind ganze Zahlen, $h, k = 0, 1, 2, \ldots$ Bei einer quadratischen Membran $a = b$ treten harmonische Oberschwingungen (Frequenz = ganzzahliges Vielfaches der Grundschwingung) und unharmonische auf. Wird für diesen Fall z. B. $\nu_{00} = 1$ gesetzt, so sind $\nu_{01} = 1{,}58$; $\nu_{02} = 2{,}24$; $\nu_{10} = 1{,}58$, aber $\nu_{11} = 2$, letzteres also eine harmonische Oberschwingung. Während bei der Grundschwingung nur der Rand eine

Knotenlinie bildet, treten bei den Oberschwingungen kompliziert verlaufende Knotenlinien auf. Für eine kreisförmige Membran, deren Eigenfrequenzen sich nicht allgemein angeben lassen, ist für die tiefsten Eigenfrequenzen der Verlauf der Knotenlinien in Abb. 2.2.2–5 skizziert. Hierbei treten keine harmonischen Oberschwingungen auf.

Platten als biegesteife Gebilde führen wie die Balken bzw. Stäbe Biegungsschwingungen aus, für die dieselben Besonderheiten wie im eindimensionalen Fall gelten. Für die Eigenfrequenzen ν_k gilt $\nu_k \sim d/r^2$ (d = Plattendicke, r = Plattenradius). Weiter hängt ν_k außer vom Elastizitätsmodul wie im Fall der Biegungsschwingungen des Balkens wegen der Zweidimensionalität der Ausbreitung auch von der Querkontraktionszahl des Plattenmateriales ab. Der Verlauf der Knotenlinien schwingender Platten ist i. allg. sehr kompliziert. Er kann durch Aufstreuen von Sand sichtbar gemacht werden, da dieser sich in den Knotenlinien sammelt, wodurch die *Chladnischen Klangfiguren* entstehen. Je komplexer das Bild der Figuren, desto höher ist die Schwingungsfrequenz der Platte.

2.2.2.3. Dreidimensionale Wellenausbreitung. Die Ausbreitung der Wellen wird mit der *dreidimensionalen Wellengleichung* (vgl. 2.3.3.) beschrieben. In die Ausbreitungsgeschwindigkeit $c_{l,3}$ einer longitudinalen Welle geht jetzt infolge der Dreidimensionalität neben dem Elastizitätsmodul E noch die Querkontraktionszahl ν ein, $c_{l,3} = \{(1 - \nu)/[(1+\nu)(1-2\nu)]\}^{1/2}(E/\varrho)^{1/2}$. Da der Faktor in der geschweiften Klammer immer größer als Eins ist, so ist die Wellenausbreitungsgeschwindigkeit in einem allseitig ausgedehnten Medium immer größer als in einem Stab des gleichen Materiales. Einige Werte sind in Tab. 2.2.2–1 angegeben. Die Fortpflanzungsgeschwindigkeit der Transversalwellen c_t ist dieselbe wie im eindimensionalen Fall, $c_t = (G/\varrho)^{1/2}$ (G = Schermodul des Mediums).

Oberflächenwellen treten bei der Wellenausbreitung an Grenzflächen auf. Ihre Eindringtiefe in das Medium beträgt nur einen Bruchteil ihrer Wellenlänge λ, für ein inkompressibles Medium z. B. ist bei einer Eindringtiefe von $\lambda/2\pi$ die Größe der Amplitude nur noch der Bruchteil 1/e des Wertes an der Grenzfläche. Die Ausbreitungsgeschwindigkeit c_{ob} der Oberflächenwellen ist etwas geringer als die der Transversalwellen. Im Fall eines inkompressiblen Mediums ist $c_{ob} = 0{,}96 \cdot c_t$. Die unterschiedlichen Ausbreitungsgeschwindigkeiten der drei Wellentypen haben eine große Bedeutung für die *Seismik*. Bei einem Erdbeben entstehen alle drei Typen. Wegen ihrer wesentlich größeren Geschwindigkeit treffen am Beobachtungsort zuerst die Longitudinalwellen als Vorläufer ein, danach die transversalen Raumwellen. Etwas später treffen die Oberflächenwellen ein, einmal wegen ihrer etwas geringeren Geschwindigkeit und zum anderen wegen des längeren Weges, den sie auf dem Kreisbogen zurückzulegen haben gegenüber der Kreissehne, entlang der sich die Raumwellen ausbreiten. Mit wachsender Entfernung vom Bebenzentrum nimmt die Bedeutung der Oberflächenwellen wegen ihrer nur zweidimensionalen Ausbreitung gegenüber den Raumwellen zu, die sich dreidimensional im Volumen verteilen, so daß erstere i. allg. den Hauptteil der beobachteten Erderschütterung bilden.

Auch räumlich begrenzte Festkörper lassen sich zu Eigenschwingungen anregen. Die Schwingungsformen sind jedoch auch bei einfacher geometrischer Gestalt der Körper (Würfel, Zylinder) sehr kompliziert. Sie lassen sich z. B. in Glaskörpern bei Betrachtung im polarisierten Licht infolge

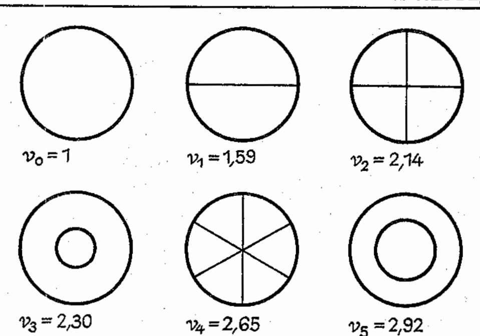


Abb. 2.2.2–5 Knotenlinien der tiefsten Eigenfrequenzen einer kreisförmigen Membran

der mit den mechanischen Spannungen verbundenen Doppelbrechung (vgl. 2.5.2.3.) sichtbar machen. Gelegentlich werden kubische oder zylindrische Körper zur Erzeugung von Ultraschall (vgl. 2.3.9.) verwendet.

2.2.2.4. Festkörper als Schallgeber, Resonanzkörper. Zur Erzeugung von *Schallwellen* (vgl. 2.3.8.) werden vorwiegend ein- und zweidimensionale *Schallgeber* verwendet. Der wichtigste eindimensionale Schallgeber ist die transversal schwingende, beidseitig eingespannte Saite (Violine, Cello, Klavier u. a.). Die Frequenzen ihrer Eigenschwingungen betragen (vgl. 2.2.2.1. und Abb. 2.2.2–2) $\nu_k = (\sigma_s/\varrho)^{1/2}(k+1)/2l$ mit $k = 0, 1, 2 \ldots$, d. h., die Oberschwingungen sind Harmonische der Grundschwingung (σ_s = Spannung, l = Länge der Saite, ϱ = Dichte des Saitenmateriales). Damit kann die Frequenz der Saite sowohl durch Änderung

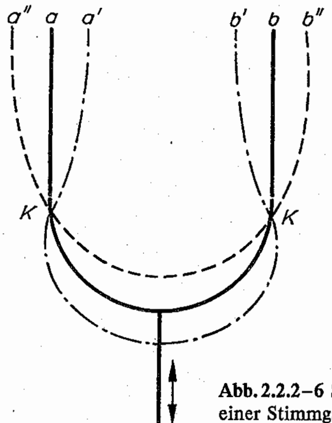


Abb. 2.2.2–6 Schwingungsformen einer Stimmgabel

der Spannkraft als auch durch Erzeugung von Schwingungsknoten durch Berühren der Saite verändert werden. Im letzteren Fall werden tiefere Frequenzen ausgelöscht. Während eine freischwingende Saite außer dem Grundton gleichzeitig eine Reihe von Obertönen liefert, erzeugt eine *Stimmgabel* einen praktisch oberschwingungsfreien Ton. Bei letzterer handelt es sich um einen U-förmig gebogenen Stab, der Biegungsschwingungen ausführt. Wie in Abb. 2.2.2–6 skizziert, schwingen die beiden Schenkel gegeneinander (jeweils die Positionen a, b; a', b'; a'', b'' entsprechen einander), während sich der zwischen den Knotenstellen K liegende Abschnitt mit dem daran befindli-

chen Stiel hebt und senkt. Durch Biegungsschwingungen von Stäben werden weiterhin die Töne beim Xylophon sowie mittels schwingender Zungen (einseitig eingespannte Stäbe) bei den Zungenpfeifen von Orgel und Harmonium, bei der Zieh- und Mundharmonika sowie bei der Spieldose erzeugt.
Wegen der relativ kleinen Oberfläche einer Saite bzw. eines Stabes ist auch die Schallabstrahlung entsprechend gering, da nur die Luft längs dieser Fläche zu Schwingungen angeregt wird. Deshalb besitzen alle Saiteninstrumente einen *Resonanzkörper* oder *Resonanzboden*, der zum Mitschwingen angeregt wird und eine wesentlich größere Schallabstrahlung als die einzelne Saite hat. Entgegen der Bezeichnung soll keine eigentliche Resonanz erzeugt werden, d. h., der Resonanzkörper soll nicht mit einer Frequenz einer Saite des Instrumentes in Resonanz geraten und bevorzugt die entsprechende Schwingung abstrahlen, sondern es wird durch geeignete Bauart versucht zu erreichen, daß alle im Frequenzbereich des Instrumentes liegenden Töne möglichst gleichmäßig verstärkt werden. Anders ist es bei der Stimmgabel. Diese wird häufig auf einen Resonanzkasten gesetzt, der so dimensioniert ist, daß die Grundschwingung der Luftsäule im Kasten dieselbe Frequenz wie die Stimmgabel hat.

2.2.3. Anelastizität

Wird bei der Verformung die Elastizitätsgrenze E überschritten (vgl. Abb. 2.2.1–1), so verhält sich der Werkstoff anelastisch, d. h., es tritt zum einen eine bleibende Verformung auf und zum anderen wird die gesamte Verformung zeitabhängig. Letzteres hat zur Folge, daß ein Körper auf eine bestimmte Beanspruchung erst mit einer zeitlichen Verzögerung reagiert.

2.2.3.1. Plastizität. Bleibt nach der Entlastung eines Festkörpers eine bleibende Verformung zurück, so ist das ein Zeichen dafür, daß während der vorangehenden Beanspruchung die *Elastizitätsgrenze* bzw. die *Fließgrenze* (vgl. 2.2.1.) überschritten war, d. h., daß plastische Verformung stattgefunden hat. Der Grenzwert der Spannung, bei der das auftritt, ist materialabhängig und kann nur experimentell bestimmt werden. Bei den meisten Werkstoffen ist es möglich, diesen Grenzwert durch einen einfach zu realisierenden Versuch, z. B. mittels eines einachsigen Zugversuches, zu ermitteln. Mit Hilfe von *Fließkriterien* kann dann auch bei komplizierten Spannungszuständen berechnet werden, wann plastische Verformung des Materiales einsetzt. Die am häufigsten benutzte und i. allg. gut zutreffende Beziehung ist das *Fließkriterium* von von Mises. Plastische Verformung tritt ein, wenn $(1/3)\ (\sigma_{xx}^2 + \sigma_{yy}^2 + \sigma_{zz}^2 - \sigma_{xx}\sigma_{yy} - \sigma_{xx}\sigma_{zz} - \sigma_{yy}\sigma_{zz}) + \tau_{yz}^2 + \tau_{zx}^2 + \tau_{xy}^2 = k^2$ gilt (k = Materialkonstante). Hencky zeigte, daß die linke Seite dieser Beziehung gleich der *Gestaltänderungsarbeit* ist, d. h. gleich der gesamten Verformungsarbeit, die notwendig ist, den durch die Spannungen σ_{xx}, σ_{yy} usw. charakterisierten Endzustand zu erreichen, abzüglich der Arbeit, die für die Volumenveränderung während der Deformation aufgebracht werden mußte. Da eine reine Volumenänderung eines Körpers nie zu dessen plastischer Verformung führen kann, ist allein der die Gestalt ändernde Anteil an der Gesamtverformung dafür maßgebend. Deshalb kann das von-Misessche Fließkriterium auch als das der *maximalen Gestaltänderungsarbeit* bezeichnet werden. Wird der Körper durch einachsigen Zug z. B. in z-Richtung belastet, d. h. ist die

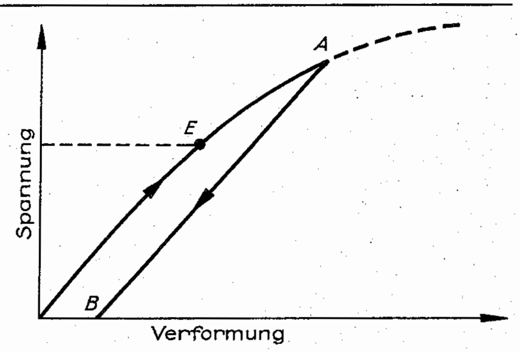


Abb. 2.2.3–1 Entlastung und Verfestigung bei anschließender Wiederbelastung

einzige von Null verschiedene Spannungskomponente σ_{zz}, so tritt plastische Verformung dann ein, wenn $\sigma_{zz}^{(F)} = \sqrt{3}\,k$ gilt. Damit kann der Materialparameter k durch diesen einfachen Versuch leicht bestimmt werden. Wird der Körper andererseits rein durch Scherung (vgl. 2.2.1.2.) belastet, d. h. sei einzig $\tau_{xy} \neq 0$, dann tritt Fließen bei $\tau_{xy}^{(F)} = k$ ein. Damit ergibt sich das Verhältnis der Fließspannung bei reinem Zug $\sigma^{(F)}$ zu der bei reiner Scherung $\tau^{(F)}$ als $\sigma^{(F)}/\tau^{(F)} = \sqrt{3}$. Andere Fließkriterien, die in der Technik ebenfalls bekannt sind, führen im Rahmen der erreichbaren Meßgenauigkeiten beim Festlegen der Elastizitäts- bzw. Fließgrenze zu mit dem obigen Kriterium vergleichbaren Ergebnissen.
Wird der Festkörper über die Elastizitätsgrenze E hinaus belastet (Abb. 2.2.3–1 als Beispiel ohne Streckgrenzenüberhöhung; vgl. 2.2.1.), so tritt kein ideales plastisches Fließen ein, d. h. eine Verformung ohne Spannungszunahme. Statt dessen wird der Festkörper verfestigt, die weitere Verformung ist mit einer, wenn auch gegenüber dem Beginn geringeren, Spannungszunahme verbunden. Wird am Punkt A entlastet, so wird der Bereich AB nahezu geradlinig durchlaufen, da nur der elastische Anteil der Verformung zurückgebildet wird. Der Punkt B zeigt die bleibende Verformung an. Bei einer erneuten Belastung wird wiederum der Bereich BA nahezu geradlinig durchlaufen und erst im durch die vorhergehende Verformung festgelegten Punkt A tritt weitere plastische Verformung auf. Beim zweiten Verformungszyklus kann also infolge der vorangegangenen *Verfestigung* bis oberhalb der durch E festgelegten Elastizitätsgrenze des „jungfräulichen" Materiales elastisch verformt werden. Das ist ein deutliches Zeichen dafür, daß das mechanische Verhalten außerhalb des elastischen Bereiches durch die Vorgeschichte des Materiales beeinflußt wird.

2.2.3.2. Rheologie. Die rheologische Beschreibung berücksichtigt die Zeitabhängigkeit des Materialverhaltens. Diese ist sowohl der Größe als auch der Art nach sehr vom Werkstoff abhängig, so daß sich hier keine allgemeingültigen Gesetzmäßigkeiten angeben lassen. Es ist deshalb notwendig, in jedem Einzelfall ganz spezielle Modelle zu entwickeln, denen nur eines gemeinsam ist, daß in den sich ergebenden mathematischen Beziehungen zwischen Spannungen und Dehnungen im Gegensatz zu den elastizitätstheoretischen Ausdrücken die Zeit jeweils explizit enthalten ist. Es lassen sich dann solche zeitabhängigen Experimente, wie der *Kriechversuch* (Verformung bei konstanter Spannung), die *Spannungsrelaxation* (Abnahme der Spannung bei konstanter Dehnung) und *mechanische Erholung* (Änderung der Dehnung nach, nicht während der Entlastung), beschrei-

ben. Solche rheologischen Eigenschaften von Festkörpern spielen in der Technik immer dann eine Rolle, wenn relativ hohe Belastungen (Leichtbau!) als Dauerbeanspruchung wirken.

2.2.4. Hydrostatik

2.2.4.1. Hydrostatischer Druck. Bei den Betrachtungen wird zunächst der Einfluß der Schwerkraft gegenüber den äußeren Drücken vernachlässigt. Wegen der leichten Verschiebbarkeit der Teilchen gegeneinander besitzen Flüssigkeiten im Gegensatz zu Festkörpern keine Gestaltelastizität. Das bedeutet, daß ein auf eine Flüssigkeit ausgeübter Druck (hydrostatischer Druck) sich durch die gesamte Flüssigkeit nach allen Seiten mit der gleichen Stärke verteilt, so daß an allen Stellen sowohl im Inneren als auch auf den Begrenzungsflächen der gleiche Druck wirkt (*Gesetz von* PASCAL). Wird z. B. ein Stempel in einem Gehäuse mit einer bestimmten Kraft in die Flüssigkeit gedrückt, so ist der Druck an allen am Umfang des Behälters angebrachten Ausflußöffnungen derselbe, wie sich durch das gleichmäßige Herausspritzen der Flüssigkeit aus allen Öffnungen zeigt. Eine wichtige Anwendung findet diese Gesetzmäßigkeit beim *hydraulischen Heber* (Abb. 2.2.4–1). Auf einen Stempel mit dem Querschnitt A_1 wirkt eine Kraft F_1 und erzeugt in der Flüssigkeit den hydrostatischen Druck $p = F_1/A_1$. Wegen der allseitigen Druckausbreitung wirkt derselbe Druck am Ort des Stempels mit dem Querschnitt A_2 und ruft dort eine Kraft $F_2 = pA_2$ hervor, die den Stempel nach oben bewegt. Wird aus dem letzten Ausdruck p eliminiert, so ist $F_2 = F_1(A_2/A_1)$ und, da $A_2 > A_1$, folgt $F_2 > F_1$, d. h., mit einer relativ kleinen Kraft F_1 kann eine entsprechend große Last F_2 gehoben werden. Wie beim Hebel (vgl. 2.1.2.) wird allerdings keine Arbeit gewonnen, sondern die Erhöhung der Kraftwirkung beim Stempel A_2 muß mit einem vergrößerten Weg beim Stempel A_1 ausgeglichen werden, so daß bei Berücksichtigung des von den beiden Stempeln zurückgelegten Weges s_1 und s_2 gilt $F_1s_1 = F_2s_2$. Eine weitere Anwendung findet das Gesetz der allseitigen Druckausbreitung beim *hydraulischen Gestänge*, das z. B. bei Bremsleitungen häufig angewendet wird (Abb. 2.2.4–2). Der Querschnitt der mit einer Flüssigkeit gefüllten Leitung bleibt i. allg. konstant, so daß wegen des konstanten Druckes auch die Kraftwirkung in jedem Querschnitt dieselbe ist. Es kann aber ohne komplizierte mechanische Vorrichtungen, wie sie bei Verwendung von Gestängen nötig wären, die Kraftrichtung in gewünschter Weise geändert werden.

Infolge der Schwerkraft wird auf eine in eine Flüssigkeit eingebrachte Fläche A (Abb. 2.2.4–3) ein zum äußeren Druck p_0 zusätzlicher Druck Δp ausgeübt, der nur von der Höhe h der Flüssigkeitssäule oberhalb der Fläche sowie von der Dichte ϱ der Flüssigkeit abhängt, $\Delta p = h\varrho g$ (g = Erdbeschleunigung). Wegen der allseitigen Druckausbreitung wirkt der Druck $p = p_0 + \Delta p$ gleichzeitig als *Seitendruck* und als *Aufdruck* in der Höhe h. Ersterer läßt sich nachweisen, wenn das Gefäß in der Entfernung h unterhalb des Flüssigkeitsspiegels eine Öffnung aufweist. Ein dort angebrachtes Manometer zeigt den Druck an. Der Aufdruck läßt sich mit Hilfe der in Abb. 2.2.4–4 angegebenen Anordnung demonstrieren. Ein Zylinder, der unten durch eine nur angelegte Scheibe verschlossen ist, wird bis zur Tiefe h in die Flüssigkeit getaucht. Infolge des Aufdruckes wird die Scheibe angepreßt und fällt erst ab, wenn die Summe aus äußerem Druck p_0 und dem Druck, der durch ein zusätzliches Gewicht G hervorgerufen wird, den Aufdruck $p = p_0 + \varrho gh$ übersteigt. Wird der durch die Schwere hervorgerufene Druck am Boden der Flüssigkeit gemessen (vgl. Abb. 2.2.4–3), so wird er als *Bodendruck* p_B bezeichnet, $p_B = p_0 + h_B\varrho g$. Die Abhängigkeit des Bodendruckes allein von der Höhe der Flüssigkeitssäule gibt Anlaß zum *hydrostatischen Paradoxon*: Die Kraftwirkung auf die Unterlage ist bei Gefäßen gleicher Grundfläche und bei gleicher Höhe der Flüssigkeitssäule unabhängig von der Gefäßform. Das hat den gleich hohen Flüssigkeitsstand in *kommunizierenden Röhren* unabhängig von Form und Neigung der Röhren zur Folge. Eine praktische Anwendung des Prinzipes der kommunizierenden Röhren ist der *Wasserstandsanzeiger* (Abb. 2.2.4–5), der bei verschiedensten Flüssigkeitsbehältern benutzt wird. Bilden wasserführende Schichten im Erdreich eine Mulde (Abb. 2.2.4–6), so entsteht durch eine Bohrung durch die wasserundurchlässige Schicht hindurch ein *artesischer Brunnen*, wie er z. B. aus der Sahara bekannt ist. Ebenfalls auf der Wirkung des hydrostatischen Druckes be-

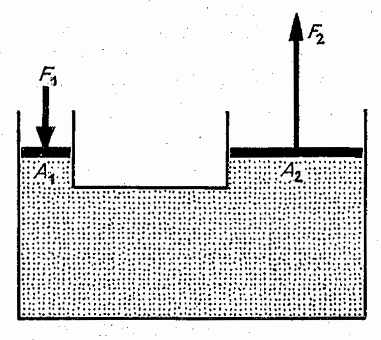


Abb. 2.2.4–1 Prinzip des hydraulischen Hebers

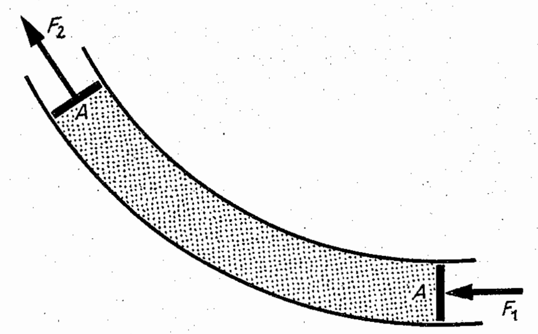


Abb. 2.2.4–2 Prinzip des hydraulischen Gestänges

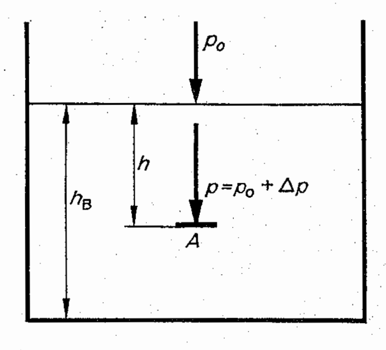


Abb. 2.2.4–3 Druck auf eine Probefläche A in Flüssigkeit

ruht der *Saugheber* (Abb. 2.2.4–7). Wird dafür gesorgt, daß $h_2 > h_1$ ist, befindet sich also die Ausflußöffnung des Schlauches ständig unterhalb der Flüssigkeitsoberfläche, so ist die Druckdifferenz $(h_2 - h_1)\varrho g > 0$, und die Flüssigkeit kann über den Rand des Behälters hinweg entleert werden. Allerdings ist ein äußerer Mindestdruck $p_0 > h_1 g$ nötig, um die Flüssigkeit bis zur Höhe h_1 zu heben, d.h., im luftleeren Raum funktioniert der Saugheber i. allg. nicht. Nur bei sehr vorsichtiger Handhabung kann unter Umständen auch im letzteren Fall durch die Kohäsionskräfte der Flüssigkeit ein Abreißen des Flüssigkeitsfadens vermieden werden.

2.2.4.2. Auftrieb. Wird ein Körper in eine Flüssigkeit eingetaucht, so entsteht infolge des unterschiedlichen hydrostatischen Druckes an Ober- und Unterseite des Körpers eine entgegen der Schwerkraft gerichtete Kraft, der Auftrieb F_A (Abb. 2.2.4–8), als Differenz der von den Drücken $p_1 = h_1\varrho g$ und $p_2 = h_2\varrho g$ herrührenden Kräfte. Es gilt bei beliebiger Form des von dem Körper verdrängten Flüssigkeitsvolumens $V_{\text{Verdr.}}$ für den Auftrieb $F_A = V_{\text{Verdr.}}\varrho_{\text{Fl}}g$, wobei ϱ_{Fl} die Dichte der Flüssigkeit ist, d. h., der Auftrieb ist gleich dem Gewicht der von dem Körper verdrängten Flüssigkeit (*archimedisches Prinzip*.)

Ein Körper sinkt in einer Flüssigkeit, wenn sein Gewicht größer als der maximal mögliche Auftrieb ist, d. h. wenn diese Bedingung auch bei völligem Eintauchen erfüllt ist. Wenn in letzterem Fall Auftrieb und Gewicht gerade gleich sind, so wird dieser Zustand als *Schweben* bezeichnet. Ist der Auftrieb bei vollständigem Eintauchen des Körpers größer als sein Gewicht, so steigt der Körper solange aus der Flüssigkeit auf, bis infolge der damit verbundenen Verminderung des verdrängten Flüssigkeitsvolumens und folglich des Auftriebes dieser gleich dem Gewicht wird. Dieser sich einstellende Gleichgewichtszustand wird als *Schwimmen* bezeichnet. Handelt es sich um homogene Körper, bei denen, also die Dichte ϱ_K in jedem Volumenelement dieselbe ist, so liefert der Vergleich der Dichten von Flüssigkeit ϱ_{Fl} und eingetauchtem Körper ϱ_K das einfachste Kriterium, um über das Verhalten des Körpers zu entscheiden: $\varrho_K > \varrho_{\text{Fl}}$ Sinken, $\varrho_K = \varrho_{\text{Fl}}$ Schweben, $\varrho_K < \varrho_{\text{Fl}}$ Schwimmen. Das gilt nicht nur für Festkörper, sondern auch für nichtmischende Flüssigkeiten und für Gase in Gasblasen, die sich nicht in der Flüssigkeit lösen. Die Flüssigkeit mit der geringeren Dichte schwimmt auf der mit der größeren (z. B. Öl auf Wasser), Gasblasen steigen in einer Flüssigkeit immer nach oben (Gasblasen in Sekt).

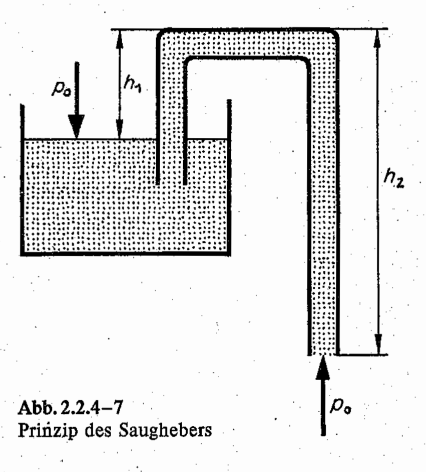


Abb. 2.2.4–7
Prinzip des Saughebers

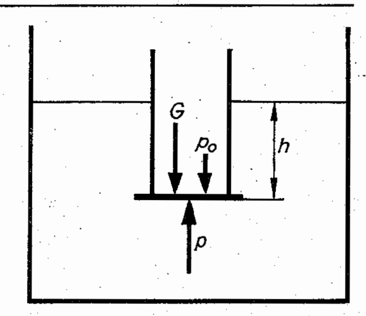
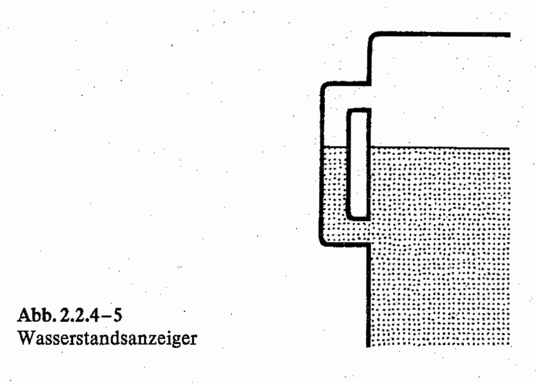
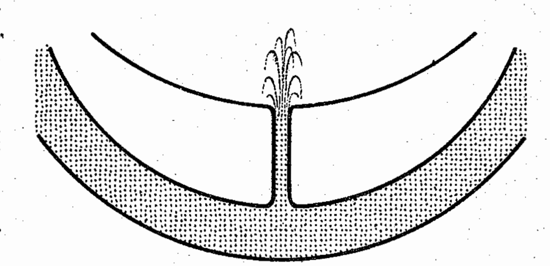


Abb. 2.2.4–4 Nachweis des Aufdruckes


Abb. 2.2.4–5
Wasserstandsanzeiger


Abb. 2.2.4–6 Artesischer Brunnen

2.2.4.3. Schwimmen. Während die Schwerkraft im Schwerpunkt S_K des Körpers angreift, ist der Angriffspunkt des Auftriebes der Schwerpunkt S_{Fl} der verdrängten Flüssigkeitsmenge, so daß im Fall des Schwimmens i. allg. beide Punkte auch bei homogenen Körpern nicht zusammenfallen. Im Gleichgewicht wirken aber Auftrieb F_A und Gewicht F_G in derselben Wirkungslinie (vgl. die jeweils linken Darstellungen in den Abb. 2.2.4–9 und –10), so daß der Körper, da beide gleich groß sind, sich in Ruhe befindet. Wird der Körper geringfügig aus der Ruhelage ausgelenkt, so tritt ein Kräftepaar auf, das eine Drehung des Körpers hervorruft. In der Abb. 2.2.4–9 bewirkt das Kräftepaar eine Drehung von der ursprünglichen Gleichgewichtslage weg, es handelt sich in diesem Fall um ein labiles Gleichgewicht. Im Gegensatz dazu bewirkt das Kräftepaar in Abb. 2.2.4–10 eine Drehung in die ursprüngliche Gleichgewichtslage zurück, es ist also ein stabiles Gleichgewicht. Entscheidend für dieses Verhalten ist die Lage des sogenannten *Metazentrums M*. Dieses ist

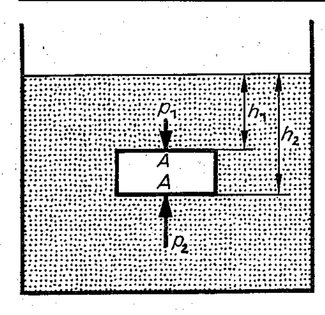


Abb. 2.2.4–8 Zur Entstehung des Auftriebes

der Schnittpunkt der Wirkungslinie des Auftriebes in der gekippten Lage mit der Verbindungslinie der Schwerpunkte S_K des Körpers und S_{Fl} der Flüssigkeit, die sich in der nichtgekippten Lage ergeben hatte (vgl. Abb. 2.2.4–9 und –10). Je höher das Metazentrum über dem Körperschwerpunkt liegt, um so größer ist die Stabilität der Gleichgewichtslage; liegt es unterhalb, handelt es sich um eine labile Schwimmlage. Durch Beschweren der Bodenfläche kann der Körperschwerpunkt so weit herabgesetzt werden, daß ein rechteckiger Körper auch aufrecht eine stabile Schwimmlage erreicht. Letzteres hat Bedeutung bei der Konstruktion und Belastung von Schiffen.

2.2.4.4. Oberflächenspannung. Auf ein Teilchen im Inneren einer Flüssigkeit werden durch die benachbarten Teilchen anziehende Kräfte ausgeübt. Da es sich um gleichartige Teilchen handelt, die gleichmäßig um das betrachtete Teilchen angeordnet sind, ist die Summe dieser Kräfte Null. Das trifft für Teilchen an der Oberfläche, deren Wirkungssphäre über die Oberfläche hinausragt, nicht mehr zu. Ein Teil der mit dem betrachteten Teilchen wechselwirkenden ist anderer Art, so daß die resultierende Kraft i. allg. nicht mehr Null ist. An einer Grenzfläche Flüssigkeit – Gas z. B. ist die Wirkung der Gasteilchen auf die Teilchen in der Oberfläche bzw. in der Nähe der Oberfläche geringer als die der benachbarten Flüssigkeitsteilchen. Letztere erzeugen deshalb eine in die Flüssigkeit hineingerichtete Kraft. Um ein Teilchen aus dem Inneren an die Oberfläche zu befördern, muß diese *Kohäsionskraft* überwunden und damit Arbeit geleistet werden, d. h., ein Teilchen an der Oberfläche besitzt potentielle Energie, die *Oberflächenenergie*. Die Arbeit, die erforderlich ist, die Oberfläche zu vergrößern, dividiert durch den Flächenzuwachs, ist die *Oberflächenspannung* σ. Ihr Wert hängt von der Art der betrachteten Substanz sowie von der Art des angrenzenden Mediums ab, weshalb sie auch, besonders in der Verbindung mit Festkörpern, als *Grenzflächenspannung* bezeichnet wird. In der Tab. 2.2.4–11 sind für einige Stoffkombinationen Oberflächenspannungen angegeben.

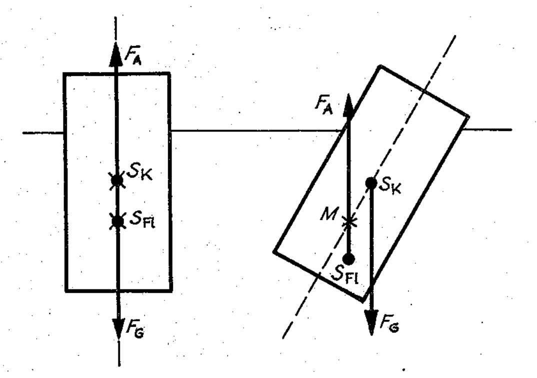


Abb. 2.2.4–9 Labile Schwimmlage

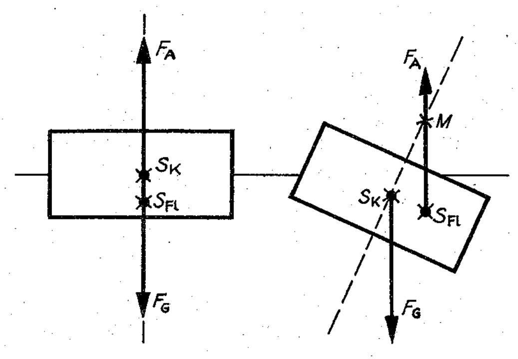


Abb. 2.2.4–10 Stabile Schwimmlage

Ist die Flüssigkeitsoberfläche nicht eben, wie bisher vorausgesetzt, sondern gekrümmt, so ruft die Kohäsionskraft einen *Krümmungsdruck* $p = \sigma(1/r_1 + 1/r_2)$ hervor, wobei r_1 und r_2 die beiden Hauptkrümmungsradien des betrachteten Flächenelementes sind (Abb. 2.2.4–12). Die Ursache ist die unterschiedliche Zahl fremder Teilchen innerhalb der Wirkungssphäre des betrachteten Teilchens in der Oberfläche. Der Krümmungsradius und damit der Krümmungsdruck sind positiv bei von unten konvexer Oberfläche, negativ bei konkaver Oberfläche. Im Fall der ebenen Oberfläche gilt für die Krümmungsradien $r_i \to \infty$, d. h., der Krümmungsdruck ist Null; handelt es sich um eine Kugeloberfläche, so ist $r_1 = r_2 = r$ und es gilt $p = 2\sigma/r$ (r = Kugelradius). Die Oberflächenspannung ermöglicht durch den Krümmungsdruck z. B. die Tatsache, daß eine Stahlnadel, die vorsichtig auf eine Flüssigkeitsoberfläche gebracht wird, nicht einsinkt,

Tab. 2.2.4–11 Oberflächenspannung einiger Stoffkombinationen bei Raumtemperatur

Kombination	Oberflächenspannung in N/m
Wasser (rein) – Luft	0,073
einmolare wässrige NaCl-Lösung – Luft	0,075
einmolare wässrige Propanollösung – Luft	0,042
Quecksilber – Luft	0,50
Olivenöl – Luft	0,032
Ethylalkohol – Luft	0,022
Terpentinöl – Luft	0,028
Benzen – Luft	0,029
Petroleum – Luft	0,026
Wasser – Olivenöl	0,020
Wasser – Petroleum	0,048
Wasser – Benzen	0,035
Quecksilber – Wasser	0,40
Quecksilber – Alkohol	0,37
Quecksilber – Olivenöl	0,34

oder daß bestimmte Insekten (Wasserläufer) auf der Wasseroberfläche ohne einzutauchen dahingleiten. Infolge des Krümmungsdruckes ist der Druck in einer großen Seifenblase geringer als in einer kleinen (umgekehrt wie bei einem Ballon mit einer Gummihaut), so daß die große

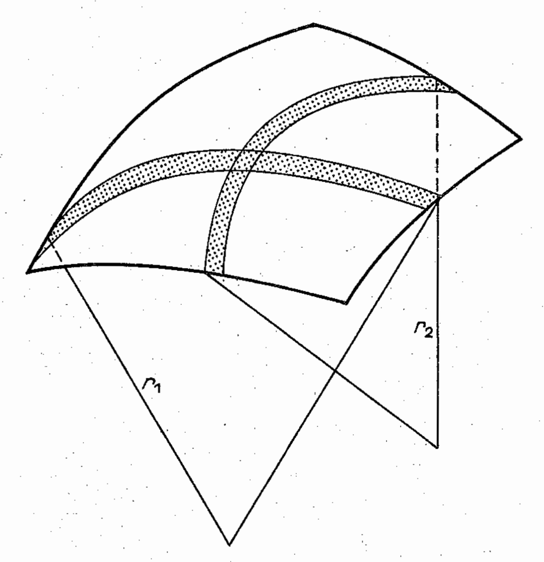


Abb. 2.2.4–12 Hauptkrümmungsradien eines Flächenelementes

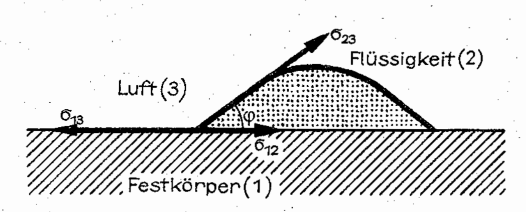


Abb. 2.2.4–13 Oberflächenspannungen am Tropfen einer benetzenden Flüssigkeit

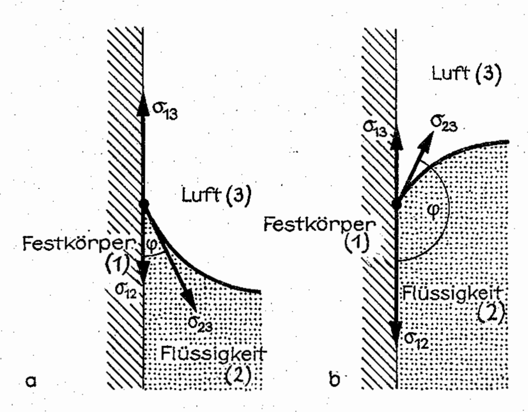


Abb. 2.2.4–14 Randwinkel φ bei (a) benetzender Flüssigkeit und (b) nichtbenetzender Flüssigkeit

Blase auf Kosten der kleineren wächst, wenn zwei Seifenblasen miteinander verbunden werden.

Stoßen in einem Punkt drei Medien zusammen, z. B. ein Flüssigkeitstropfen auf einen Festkörper in Luft (Abb. 2.2.4–13), so stellt sich dort der *Randwinkel* φ derart ein, daß Kräftegleichgewicht herrscht, $\sigma_{13} = \sigma_{12} + \sigma_{23}\cos\varphi$, wobei σ_{ij} jeweils die Oberflächenspannung für die Kombination der Medien i und j ist. Ist φ ein spitzer Winkel, so wird die Flüssigkeit bezüglich der Festkörperoberfläche als benetzend bezeichnet, im Fall eines stumpfen Winkels als nichtbenetzend. Für $\sigma_{13} - \sigma_{12} \geqq \sigma_{23}$ ist der Randwinkel stets Null, die Flüssigkeit ist in diesem Fall vollständig benetzend, ein Tropfen kann sich nicht bilden, sondern die ganze Oberfläche ist mit Flüssigkeit überzogen. Analoge Verhältnisse stellen sich ein, wenn ein Flüssigkeitstropfen auf einer anderen Flüssigkeit schwimmt. Bei Olivenöl auf Wasser z. B. kann sich ein Gleichgewicht zwischen den drei Oberflächenspannungen nicht einstellen (vgl. Werte der Tab. 2.2.4–11), so daß Öl auf Wasser immer eine dünne Haut bildet. Grenzt eine Flüssigkeit an eine feste Oberfläche, so sind die Verhältnisse wie beim Tropfen auf einer Unterlage. Der Randwinkel φ stellt sich wieder entsprechend dem Gleichgewicht zwischen den drei Oberflächenspannungen ein. Wenn z. B. Wasser an Glas angrenzt, ist es eine benetzende Flüssigkeit (spitzer Randwinkel, Abb. 2.2.4–14a), während Quecksilber an Glasoberflächen eine nichtbenetzende Flüssigkeit ist (stumpfer Randwinkel, Abb. 2.2.4–14b).

2.2.4.5. Kapillarität. Wird ein enges Rohr (Kapillare) in eine Flüssigkeit getaucht, so bildet sich entsprechend dem Gleichgewicht der Oberflächenspannungen (vgl. 2.2.4.4.) der Randwinkel heraus. Bei spitzem Randwinkel bildet der Meniskus eine von unten konkave Fläche, so daß zusätzlich zum Kohäsionsdruck ein nach oben wirkender Krümmungsdruck auftritt, der eine Hebung des Flüssigkeitsspiegels bewirkt. Im Fall einer nichtbenetzenden Flüssigkeit (stumpfer Randwinkel, der Meniskus bildet von unten eine konvexe Fläche) ruft der Krümmungsdruck eine *Kapillardepression* hervor. Es gilt für die Kapillarhöhe (Steighöhe) $h = 2\sigma\cos\varphi/(r\varrho g)$ (r = Radius der Kapillare, ϱ = Dichte der Flüssigkeit). Bei einem Randwinkel $\varphi \leqq 90°$ ist $h \geqq 0$, das bedeutet Hebung, bei $\varphi > 90°$ ist $h < 0$, es tritt Depression auf. Die Kapillardepression ist z. B. bei der exakten Ablesung eines Quecksilberbarometers zu beachten.

2.2.5. Hydrodynamik

Bei einer bewegten Flüssigkeit interessiert neben der Druckverteilung und der Massendichte $\varrho(\boldsymbol{r}, t)$ die *Geschwindigkeitsverteilung* $\boldsymbol{v}(\boldsymbol{r}, t)$ in der Flüssigkeit oder die *Stromdichte* $\boldsymbol{j}(\boldsymbol{r}, t) = \varrho(\boldsymbol{r}, t)\,\boldsymbol{v}(\boldsymbol{r}, t)$. Betrachtet man ein Volumenelement ΔV in der Umgebung des Ortes $\boldsymbol{r}$, dann befindet sich in ihm z. Z. t die Masse $\varrho(\boldsymbol{r}, t)\,\Delta V$. $\boldsymbol{v}(\boldsymbol{r}, t)$ gibt an, mit welcher Geschwindigkeit sich die Flüssigkeit in einem Volumenelement am Ort $\boldsymbol{r}$ zur Zeit t bewegt. Zu verschiedenen Zeiten befinden sich verschiedene Flüssigkeitsteile in diesem Volumenelement, d. h., es wird eine andere Beschreibungsweise (die Feldbeschreibung) als in der Punktmechanik verwendet, wo die Bahnkurve jeder einzelnen Masse verfolgt wird. Bei *stationären Strömungen* hängt $\boldsymbol{v}$ nicht von t ab. Das Geschwindigkeitsfeld $\boldsymbol{v}(\boldsymbol{r})$ kann man durch Stromlinien veranschaulichen (vgl. 2.1.3.6.). Betrachtet man ein Flächenelement ΔA, dessen Flächennormale parallel zur Geschwindigkeit $\boldsymbol{v}(\boldsymbol{r}, t)$ liegt, dann tritt in der

Zeit Δt diejenige Flüssigkeit durch ΔA hindurch, die sich in einem Volumenelement $\Delta A v(r, t)\Delta t$ befindet. Sie hat die Masse $\varrho(r, t)\Delta A v(r, t)\Delta t$. Die Massenstromdichte, definiert als Masse pro Fläche und pro Zeit, ergibt sich daraus, wenn durch ΔA und Δt dividiert wird. Die Richtung des Vektors der Massenstromdichte ist mit der der Geschwindigkeit identisch, so daß sich der Zusammenhang $j(r, t) = \varrho(r, t)\, v(r, t)$ ergibt.

Im folgenden wird immer angenommen, daß die Flüssigkeit *inkompressibel* (nicht zusammendrückbar) sei, d.h., daß ihre Dichte räumlich konstant ist. Die *Kontinuitätsgleichung* als Ausdruck der Massenerhaltung vereinfacht sich dann zu der Aussage, daß die Stromdichte quellenfrei ist (vgl. 2.1.3.6.), daß im Stromlinienbild im Innern der Flüssigkeit keine Stromlinien beginnen oder enden. Betrachtet man die Strömung in Rohren variierenden Querschnittes A, dann folgt daraus $v_1A_1 = v_2A_2$, wobei v_1 und v_2 die Strömungsgeschwindigkeiten bei A_1 und A_2 sind.

Die zeitliche Änderung der Geschwindigkeit eines Flüssigkeitselementes wird durch die *Newtonsche Bewegungsgleichung* (vgl. 2.1.3.1.) bestimmt. Sie erhält jedoch wegen der oben erläuterten Beschreibung durch ein Geschwindigkeitsfeld eine andere Gestalt. Als Kräfte treten äußere Kräfte, z. B. die Schwerkraft, Kräfte infolge von Druckdifferenzen in der Flüssigkeit und Reibungskräfte auf.

Die Reibungskraft F_R zwischen zwei sich relativ zueinander bewegenden benachbarten Flüssigkeitsschichten bewirkt, daß die schnellere Schicht verlangsamt, die langsamere beschleunigt wird. Für diese Kraft gilt $F_R = A\eta\Delta v/\Delta h$ (A = Fläche der aneinander vorbeigleitenden Schichten, η = Viskosität der Flüssigkeit, Δv = Geschwindigkeitsdifferenz, Δh = Abstand der Schichten) (Abb. 2.2.5–1). Damit wird die *tangentiale Reibungsspannung* τ_R in der Flüssigkeit $\tau_R = \eta\, dv/dh$ für infinitesimal benachbarte Schichten (dv/dh = Gradient der Geschwindigkeit senkrecht zur Strömungsrichtung). In Tab. 2.2.5–2 sind Viskositäten für einige Medien angegeben. Am Beispiel des Wassers ist zu erkennen, daß die Werte stark von der Temperatur abhängen.

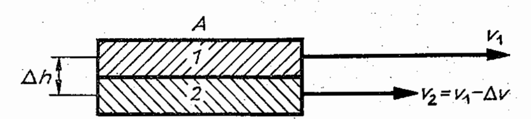


Abb. 2.2.5–1 Wechselwirkung zweier aneinander vorbeigleitender Flüssigkeitsschichten

Tab. 2.2.5–2 Viskosität flüssiger und gasförmiger Medien

Medium	Viskosität in mPa·s	Temperatur in °C
Wasser	1,81	0
	1,01	20
	0,29	100
Alkohol	1,25	20
Ether	0,26	20
Benzen	0,63	20
Quecksilber	1,59	20
Rizinusöl	969	20
Luft	0,017	0
Argon	0,021	0
Wasserstoff	0,009	0

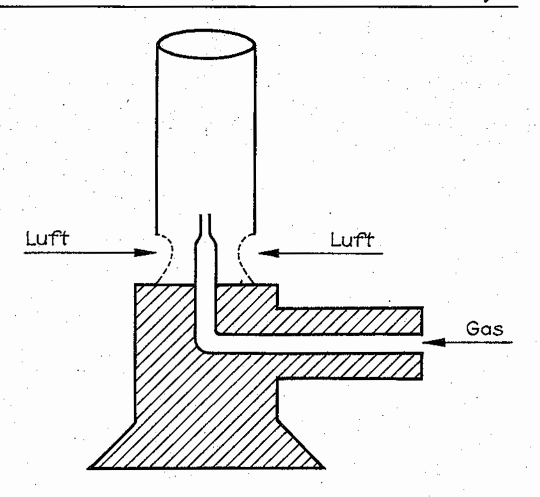


Abb. 2.2.5–3 Bunsenbrenner

2.2.5.1. Reibungsfreie Flüssigkeit. Eine viele praktisch wichtige Fälle gut beschreibende Idealisierung ist die reibungsfreie Flüssigkeit. In diesem Fall wird die tangentiale Reibungsspannung vernachlässigt. Das ist immer dann gerechtfertigt, wenn die *Viskosität* η und der Geschwindigkeitsgradient dv/dh (vgl. 2.2.5.) ausreichend klein sind. Die zweite Bedingung ist insbesondere dann zu beachten, wenn Grenzschichten behandelt werden, da dann häufig zwar die erste Bedingung zutrifft, aber nicht die zweite, so daß die Idealisierung nicht erlaubt ist.

Für die reibungsfreie Strömung folgt aus der Bewegungsgleichung die Erhaltung der Energie. Diese Aussage lautet für eine wirbelfreie Strömung $p + \varrho gh + (\varrho/2)v^2 = \text{konst.}$ und wird als *Bernoullische Gleichung* bezeichnet (p = statischer Druck, ϱgh = Schweredruck, $(\varrho/2)v^2$ = Staudruck, ϱ = Dichte der Flüssigkeit, v = Geschwindigkeit der Flüssigkeit, h = Höhe des betrachteten Punktes bezüglich eines geeignet gewählten Nullpunktes). Kontinuitätsgleichung und Bernoullische Gleichung reichen aus, um die wichtigsten Fragen bei der stationären, wirbelfreien Strömung einer inkompressiblen, reibungsfreien Flüssigkeit zu beantworten. Strömt aus einem Behälter mit dem Querschnitt A_2 aus einer kleinen Öffnung des Querschnittes A_1 ($A_1 \ll A_2$ als auch $v_2 \ll v_1$) Flüssigkeit aus, so folgt sofort das *Torricellische Theorem* für die Ausflußgeschwindigkeit $v_1 = (2\,gh)^{1/2}$ durch Anwendung der Bernoullischen Gleichung auf die beiden Querschnitte A_1 und A_2, wenn man $p_1 = p_2 = p_0$ setzt. Die Ausflußgeschwindigkeit ist so groß, als wenn die Flüssigkeitsteilchen die Höhe h vom oberen Flüssigkeitsspiegel zum unteren durchfallen hätten. Wird eine ebene Rohrströmung betrachtet ($h = 0$ für alle Querschnitte), so sagt die Bernoullische Gleichung aus, daß der statische Druck um so kleiner ist, je größer die Geschwindigkeit in dem jeweiligen Querschnitt ist. Damit wird z. B. das *hydrodynamische Paradoxon* erklärt. Ein Medium mit $p_1 \approx p_0$ strömt mit der Geschwindigkeit v_1 aus einem Rohr gegen eine Scheibe. Entgegen der Erwartung wird die Scheibe nicht vom Rohr weggestoßen, sondern an die Rohröffnung herangezogen. Da der Rohrquerschnitt größer ist als der Querschnitt zwischen Rohr und Scheibe, so ist infolge der Kontinuitätsgleichung (vgl. 2.2.5.) die Geschwindigkeit parallel zur Scheibe

$v_2 > v_1$, d. h., der statische Druck an der Scheibe ist kleiner als der äußere Luftdruck p_L, so daß letzterer die Scheibe gegen die Rohröffnung preßt. Die Wirkung von *Wasserstrahlpumpe* und *Bunsenbrenner* (Abb. 2.2.5–3) läßt sich ebenfalls mit Hilfe der Bernoullischen Gleichung erklären. In beiden Fällen wird durch eine Verengung des Querschnittes des strömenden Mediums (Wasser bzw. Gas) die Strömungsgeschwindigkeit erhöht und damit der statische Druck vermindert, so daß jeweils von der Seite Luft angesaugt wird. Befindet sich ein Gas unter einem Druck p_1 in einem größeren Gefäß und strömt durch eine kleine Öffnung gegen den Außendruck p_2 aus, so ist die Ausströmgeschwindigkeit $v_A = [2(p_1 - p_2)/\varrho]^{1/2}$ und es gilt das Gesetz von BUNSEN, daß sich bei gleicher Druckdifferenz die Ausströmgeschwindigkeit zweier Gase umgekehrt wie die Wurzeln aus ihren Dichten verhalten. Handelt es sich um die Brennkammer einer Rakete, so ist der Rückstoß $F_{\text{Rück}} = v_A\, dm/dt = v_A(\varrho A_0 v_A) = 2A_0(p_1 - p_2)$ (A_0 = Fläche der Austrittsöffnung, ϱ = Dichte des Gases, dm/dt = je Zeiteinheit austretende Gasmasse).

Die für eine wirbelige Strömung charakteristischen *Wirbel* kann man im einfachsten Fall daran erkennen, daß im Stromlinienbild geschlossene Linien auftreten. Die *Wirbelstärke* ist durch das Linienintegral $I_W = \oint d\boldsymbol{r}\boldsymbol{v}(\boldsymbol{r})$ über einen geschlossenen Weg definiert (vgl. 2.1.3.6.), das mittels des *Gaußschen Satzes* (vgl. 13.4.6.3.) in ein Flächenintegral $I_W = \int d\boldsymbol{a} \operatorname{rot} \boldsymbol{v}$ überführt werden kann. Für eine mit konstanter Winkelgeschwindigkeit ω rotierende Flüssigkeit ergibt sich $I_W = 2\pi\omega r^2$, wenn man über einen Kreis vom Radius r um den Wirbelmittelpunkt integriert. Eine von r unabhängige Wirbelstärke erhält man, wenn der Betrag der Geschwindigkeit proportional zu $1/r$ abfällt.

Für ein reibungsfreies Medium gelten die *Helmholtzschen Wirbelsätze*. 1. Die Wirbelintensität ist eine unveränderliche und unzerstörbare Eigenschaft eines Wirbels. 2. Ein Wirbel kann niemals innerhalb eines Mediums enden. Aus dem ersten Satz folgt z. B., daß in einer reibungsfreien Flüssigkeit keine Wirbel entstehen können. Sie bilden sich deshalb stets hinter festen Körpern infolge von Reibungsvorgängen in der Grenzschicht. Aus dem ersten Satz folgt auch die an Rauchringen leicht zu beobachtende Tatsache, daß sich die Wirbelgeschwindigkeit mit Zunahme des Durchmessers des Rauchringes vergrößert, da letzteres mit einer Abnahme des Wirbelquerschnittes verbunden ist. Aus dem zweiten Satz folgt, daß Wirbel entweder als Wirbelringe (z. B. Rauchringe) in sich geschlossen sind oder daß sie an Flüssigkeitsoberflächen enden. Wirbelringe besitzen eine relativ große Steifigkeit, so daß sie Hindernisse umwerfen bzw. selbst als verhältnismäßig starke Hindernisse wirken können, z. B. beim Schwimmen im Wasser.

2.2.5.2. Viskose Flüssigkeit. Der Einfluß der Reibungskräfte auf die Strömung ist nicht mehr zu vernachlässigen. Überwiegt deren Einfluß den der Trägheitskräfte, so liegt eine *laminare Strömung* vor, d. h., alle Flüssigkeitsteilchen beschreiben Bahnen, die stets parallel zueinander verlaufen. Nimmt die Größe der Trägheitskräfte zu, i. allg. durch eine Erhöhung der Geschwindigkeit bei sonst gleichbleibenden Verhältnissen, so schlägt die Bewegung plötzlich in eine *turbulente Strömung* um, bei der eine Vermischung benachbarter Flüssigkeitselemente eintritt. Aus dem Verhältnis von Trägheits- zu Reibungskräften läßt sich mittels Dichte ϱ, Viskosität η und Strömungsgeschwindigkeit v des Mediums sowie einer für das Problem charakteristischen Länge l ein dimensionsloser Ausdruck, die *Reynoldssche Zahl* $Re = \varrho v l/\eta$, ableiten. Große Werte von Re haben *Turbulenz*, kleine *Laminarität* zur Folge. Der kritische Wert von Re, bei dem der Umschlag einsetzt, ist noch stark von den spezifischen Besonderheiten der Strömung abhängig. Beispielsweise ist für eine Rohrströmung mit dem Rohrradius $r = l$ der kritische Wert $Re_{\text{krit}} = 1160$. Die Reynoldssche Zahl hat große Bedeutung bei der Übertragung der Ergebnisse von Modellversuchen im Strömungs- oder Windkanal auf reale Verhältnisse. Das *Ähnlichkeitsprinzip* ist nur gewahrt, wenn Re in allen zu vergleichenden Ausführungen denselben Wert hat. Wird ein Modellversuch in demselben Medium an einem verkleinerten Modell durchgeführt, so muß die Strömungsgeschwindigkeit in demselben Verhältnis erhöht werden, damit die Ähnlichkeit gewahrt bleibt. Wird im Modellversuch ein anderes Medium als im Original ausgewählt, so ist das Verhältnis der *kinematischen Zähigkeiten* $\nu = \eta/\varrho$ beider Medien maßgebend. Es gilt z. B. für Luft und Wasser $\nu_{\text{Luft}} = 14 \cdot \nu_{\text{Wasser}}$, d. h., wird der Modellversuch im Wasser ausgeführt und soll das Ergebnis auf das Original in Luft übertragen werden, so können die Abmessungen verringert werden, ohne daß die Strömungsgeschwindigkeit erhöht werden muß. Für eine *laminare Rohrströmung* gilt für das je Zeiteinheit durch ein Rohr der Länge l strömende Flüssigkeitsvolumen V/t das *Hagen-Poiseuillesche Gesetz* $V/t = \pi r^4 \Delta p/8\eta l$, ($r$ = Rohrradius, Δp = Druckdifferenz längs der Länge l, η = Viskosität der Flüssigkeit). Damit läßt sich auch der Reibungswiderstand F_{H-P}, den das Rohr der Länge l der Strömung entgegensetzt, angeben, $F_{H-P} = 8\pi\eta l v$ (v = mittlere Geschwindigkeit der Strömung).

Befindet sich ein Körper in einem flüssigen oder gasförmigen Medium und findet eine Bewegung mit der Relativgeschwindigkeit v statt, so tritt eine Widerstandskraft F_W auf, für die i. allg. $F_W = c\varrho v^2 A/2$ geschrieben werden kann (A = Fläche des bewegten bzw. angeströmten Körpers, ϱ = Dichte des Mediums, c = Widerstandszahl). Die *Widerstandszahl* ist eine Funktion der Körpergeometrie und der Reynoldsschen Zahl. Bei laminarer Strömung gilt z. B. für eine Kugel mit $l = r$ (Kugelradius) für die Widerstandszahl $c_K = 12/Re$, was zum *Stokesschen Gesetz* für die Reibungskraft $F_{St} = 6\pi\eta r v$ führt. Die Proportionalität $F_W \sim v$ zeigt an, daß es sich um Strömungsverhältnisse handelt, bei der

Tab. 2.2.5–4 Widerstandszahlen für einige angeströmte Körper

Körperform	Widerstandszahl c
Rechteckplatte a/b	
1	1,10
2	1,15
4	1,19
10	1,29
18	1,40
∞	2,01
Kreisplatte	1,11
offene Halbkugeln	
Anströmung in die Öffnung	1,33
Anströmung an die Wölbung	0,34
Kugel	
$Re > 1{,}5 \cdot 10^5$	0,1
$Re < 1{,}5 \cdot 10^5$	0,47

die Reibungskräfte überwiegen. Setzt Turbulenz ein, so dominieren die Trägheitskräfte als Strömungswiderstand. Hier ist c meist nur noch eine Funktion der Körpergeometrie und unabhängig von Re, so daß jetzt die Proportionalität $F_w \sim v^2$ gilt. In Tab. 2.2.5–4 sind für einige einfache geometrische Formen die Widerstandszahlen c angegeben.

2.3. Wellenlehre

Die Ausbreitung von Wellen (oder Feldern) wird in vielen Gebieten der Physik beschrieben. Bekannt sind Seilwellen, Wasserwellen, Schallwellen, elektromagnetische Wellen u. a. Trotz der Verschiedenheit dieser Phänomene treten eine Reihe von Gemeinsamkeiten auf.

Ein *Feld* wird in der Physik durch eine Funktion beschrieben, die jedem Raumpunkt zu jedem Zeitpunkt einen Wert einer physikalischen Größe zuordnet. Bei *skalaren Feldern*, wie Massendichte $\varrho(\boldsymbol{r}, t)$ oder Temperaturverlauf $T(\boldsymbol{r}, t)$, wird jedem Raumpunkt ein Zahlenwert zugeordnet. Durch *Vektorfelder*, wie die elektromagnetischen Felder $\boldsymbol{E}(\boldsymbol{r}, t)$ und $\boldsymbol{B}(\boldsymbol{r}, t)$ oder ein Geschwindigkeitsfeld $\boldsymbol{v}(\boldsymbol{r}, t)$, wird in jedem Raumpunkt ein Vektor angegeben. Hängen die Felder nicht von der Zeit ab, werden sie als statische oder stationäre Felder, im zeitabhängigen Fall als Schwingungen, Wellen oder Wellenpakete bezeichnet.

2.3.1. Rein periodische Wellen

Eine zeitlich und räumlich rein periodische Abhängigkeit ist ein möglicher Verlauf eines Feldes. Die Ausbreitung eines solchen Feldes ist besonders einfach zu übersehen, daher werden insbesondere solche Wellen untersucht.

Bei der mathematischen Beschreibung einer Welle und deren charakteristischen Größen wird zunächst von Einflüssen der Berandung, z. B. Reflexion, abgesehen, das betrachtete Gebiet sei also beliebig weit ausgedehnt.

2.3.1.1. Seilwellen. Wird das Ende eines Seiles auf und ab bewegt, dann breitet sich eine Welle aus. Bei einer rein periodischen Welle ist die Auslenkung durch eine einzige Cosinusfunktion (oder Sinusfunktion) beschreibbar. Bei einer Momentaufnahme zum Zeitpunkt t_0 (Abb. 2.3.1–1) ist die Auslenkung u dann durch $u(x, t_0) = u_0 \cos(kx + \varphi)$ gegeben. u_0 ist die *Amplitude*, die maximale Auslenkung. Das Argument der Cosinusfunktion $\alpha = kx + \varphi$ wird als *Phase* der Welle bezeichnet. Eine Erhöhung der Phase um ein Vielfaches von 2π beeinflußt den Funktionswert nicht, $\cos\alpha = \cos(\alpha + 2\pi)$. Bei der in Abb. 2.3.1–1 betrachteten Funktion ändert sich die Phase räumlich. Die Strecke, innerhalb der sie sich um 2π erhöht, nennt man die *Wellenlänge* λ. Sie ist durch $k(x + \lambda) = kx + 2\pi$ bestimmt, hängt also mit der *Wellenzahl* k über $k = 2\pi/\lambda$ zusammen. Die *Phasenkonstante* φ verschiebt die Funktion u als ganzes.

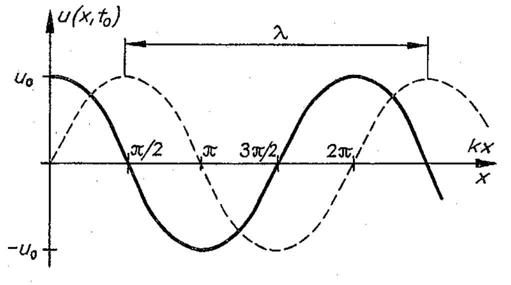


Abb. 2.3.1–1 Räumlich rein periodische Welle für $\varphi = 0$ (——) bzw. $\varphi = -\pi/2$ (- - -)

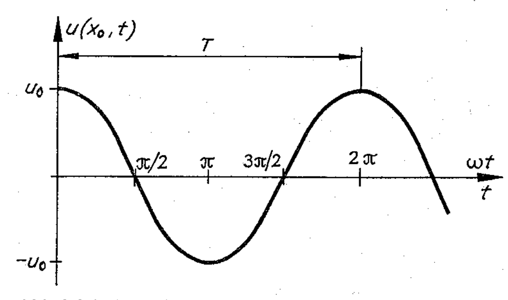


Abb. 2.3.1–2 Zeitlich rein periodische Welle, $\varphi = 0$

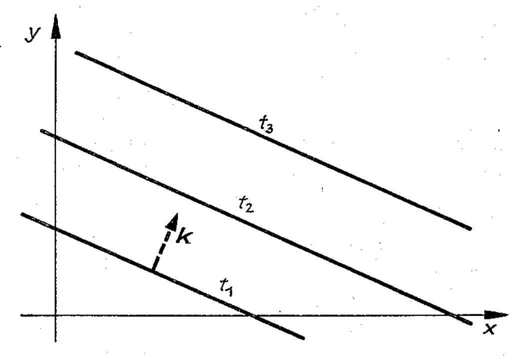


Abb. 2.3.1–3 Flächen konstanter Phase einer ebenen Welle für verschiedene Zeiten $t_1 < t_2 < t_3$

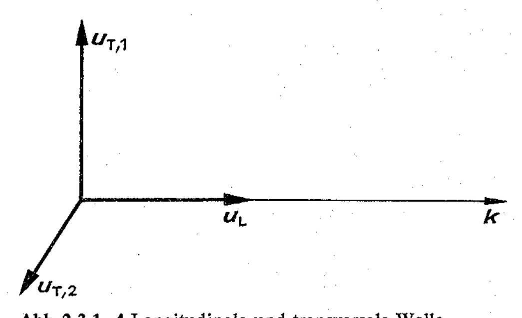


Abb. 2.3.1–4 Longitudinale und transversale Welle

Wird der Zeitablauf der Bewegung eines Punktes x_0 des Seiles verfolgt, so wird bei einer rein periodischen Bewegung $u(x_0, t) = \cos(\omega t + \varphi)$ beobachtet (Abb. 2.3.1–2). φ ist die Anfangsphase zum Zeitpunkt $t = 0$. Als *Schwingungsdauer* T bezeichnet man die Zeit, innerhalb der sich die Phase um 2π erhöht. Sie ist durch $\omega T = 2\pi$, $\omega = 2\pi f = 2\pi/T$ gegeben. Die *Kreisfrequenz* ω gibt die Änderung der Phase je Zeit an, die *Frequenz* f die Anzahl der Schwingungen je Zeit. Die Frequenz $f = 1\,\mathrm{s}^{-1}$ entspricht der Kreisfrequenz $\omega = 2\pi\,\mathrm{rad\,s}^{-1}$. Die Frequenz wird in Hertz angegeben ($1\,\mathrm{Hz} = 1\,\mathrm{s}^{-1}$). Die Frequenz $f = 1\,\mathrm{Hz}$ entspricht einer Schwingung je Sekunde, bei $f = 1\,\mathrm{kHz}$, $1\,\mathrm{MHz}$ bzw. $1\,\mathrm{GHz}$ erfolgen 10^3, 10^6 bzw. 10^9 Schwingungen je Sekunde.

Die zeitliche und räumliche Abhängigkeit der rein periodischen Seilwelle ist durch

$u(x, t) = u_0 \cos(kx - \omega t + \varphi)$, $k = 2\pi/\lambda$, $\omega = 2\pi f = 2\pi/T$

gegeben. Die in Abb. 2.3.1–1 zu einem Zeitpunkt dargestellte Welle verändert sich im Zeitablauf. Ein Punkt mit einem festen Wert der Phase, $kx - \omega t + \varphi = \text{konst.}$, verschiebt sich mit der *Phasengeschwindigkeit* $v_{ph} = \omega/k = f\lambda$ nach rechts.

2.3.1.2. Ebene Wellen. Bei einer räumlichen Wellenausbreitung lautet eine rein periodische Welle $u(\boldsymbol{r}, t) = u_0 \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)$. Legt man die x-Achse in Richtung des *Wellenvektors* $\boldsymbol{k} = k\boldsymbol{e}_x$, dann ist $\boldsymbol{k} \cdot \boldsymbol{r} = kx$, und auf der x-Achse hat man ein Wellenbild wie bei einer Seilwelle. $u(\boldsymbol{r}, t)$ ist jedoch im gesamten Raum definiert. Es gibt Flächen konstanter Phase, $\boldsymbol{k} \cdot \boldsymbol{r} - \omega t = \text{konst.}$ Da dies die Gleichung einer Ebene ist (Abb. 2.3.1–3), spricht man von einer *ebenen Welle*. Die *Richtung des Wellenvektors* $\boldsymbol{k}$ gibt die Normalenrichtung dieser Ebene an, im Zeitablauf verschiebt sich die Ebene konstanter Phase in Richtung von $\boldsymbol{k}$. Der *Betrag des Wellenvektors*, die Wellenzahl $k = 2\pi/\lambda$, bestimmt die Wellenlänge.

Da das Rechnen mit Exponentialfunktionen einfacher ist als das mit Cosinusfunktionen, verwendet man oft die *komplexe Schreibweise* für die ebene Welle:

$u = A\,e^{i(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t)}$, $A = a\,e^{i\varphi}$,
$\mathrm{Re}\,u = a \cos(\boldsymbol{k} \cdot \boldsymbol{r} - \omega t + \varphi)$.

Der Realteil der komplexen Funktion u beschreibt dann die physikalische Größe.

2.3.1.3. Polarisation. Ist die Wellenfunktion ein Vektorfeld $\boldsymbol{u}$, dann treten bei einer ebenen Welle zwei Vektoren auf, der Wellenvektor $\boldsymbol{k}$ und der Amplitudenvektor $\boldsymbol{u}_0$. Liegen beide Vektoren parallel, dann spricht man von einer *longitudinalen Welle*, stehen sie senkrecht zueinander, von einer *transversalen Welle* (Abb. 2.3.1–4). Es gibt zwei voneinander linear unabhängige transversale Wellen.

Seilwellen sind transversale Wellen. Ebenso sind elektromagnetische Wellen, d.h. auch Lichtwellen, transversal, wobei die Vektoren von elektrischem und magnetischem Feld senkrecht zueinander stehen. Bei Gitterschwingungen in Festkörpern treten sowohl longitudinale wie auch transversale Wellen auf.

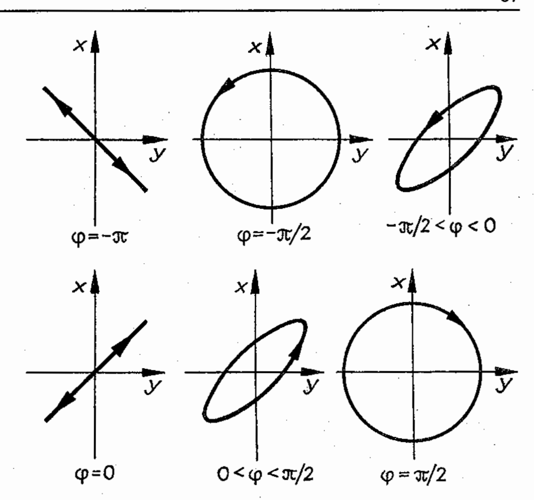


Abb. 2.3.1–6 Überlagerung zweier linear polarisierter Wellen bei verschiedenem Phasenunterschied φ mit Blick in z-Richtung

Abb. 2.3.1–5 zeigt zwei *linear polarisierte* Seilwellen verschiedener Polarisationsrichtung, $\boldsymbol{u}_{T1} = u_0\boldsymbol{e}_x$, $\boldsymbol{u}_{T2} = u_0\boldsymbol{e}_y$. Eine *zirkular polarisierte* Welle ergibt sich durch Überlagerung $\boldsymbol{u}(z, t) = u_0[\boldsymbol{e}_x \cos(kz - \omega t) + \boldsymbol{e}_y \cos(kz - \omega t + \varphi)]$ dieser linear polarisierten Wellen. Je nach dem Phasenunterschied beschreibt der Endpunkt des Vektors $\boldsymbol{u}(0, t)$ einen Kreis, eine Ellipse oder eine Gerade. Beispiele:

$\varphi = 0$: $\boldsymbol{u}(0, t) = u_0(\boldsymbol{e}_x + \boldsymbol{e}_y) \cos \omega t$, linear,

$\varphi = \pi/2$: $\boldsymbol{u}(0, t) = u_0(\boldsymbol{e}_x \cos \omega t + \boldsymbol{e}_y \sin \omega t)$, rechts zirkular,

$\varphi = -\pi/2$: $\boldsymbol{u}(0, t) = u_0(\boldsymbol{e}_x \cos \omega t - \boldsymbol{e}_y \sin \omega t)$ links zirkular.

Abb. 2.3.1–6 veranschaulicht den Zeitablauf dieser Vektoren.

2.3.1.4. Kugelwellen. Von einer punktförmigen Quelle laufen Wellen $u(r, t) = u_0 (1/r)\, e^{i(kr - \omega t)}$ radial nach außen. Man

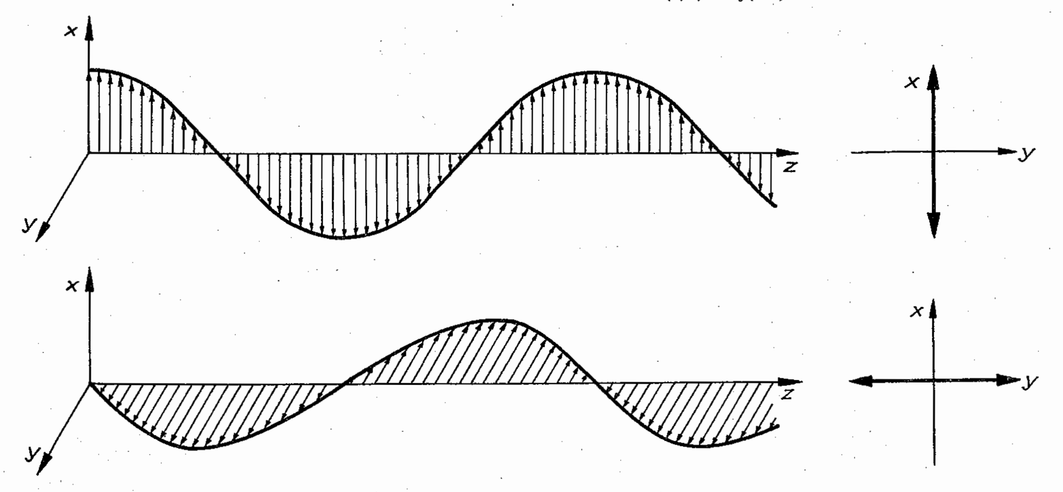


Abb. 2.3.1–5 In x- bzw. y-Richtung linear polarisierte Welle, daneben das Schwingungsbild in der Ebene $z = 0$

erhält Kreise (bei Wasserwellen) oder Kugeln, $u(r, t_0) = \text{konst.}$ für $r = \text{konst.}$, als Flächen konstanter Phase. Die Amplitude fällt mit $1/r$ ab, so daß das Produkt von u^2 *im Abstand* R mit dem Volumen einer Kugelschicht $4\pi R^2 \Delta R$ unabhängig von der Wahl von R ist. In einem kleinen Raumbereich um einen Punkt r_0 (Ausdehnung klein gegen r_0) kann eine Kugelwelle durch eine ebene Welle $(u_0/r)\mathrm{e}^{ikr} \approx (u_0/r_0)\,\mathrm{e}^{ikr_0}\mathrm{e}^{ik(r-r_0)}$ angenähert werden. Deren Wellenvektor $k = kr_0/r_0$ zeigt radial nach außen.

2.3.1.5. Stehende Wellen. Überlagern sich zwei Wellen entgegengesetzter Ausbreitungsrichtung (z. B. einfallende und reflektierte Welle) und gleicher Amplitude, dann entsteht eine stehende Welle $u(x, t) = u_0 \cos(\omega t - kx) + u_0 \cos(\omega t + kx) = 2u_0 \cos \omega t \cos kx$. Die Knoten $(u = 0)$ der Wellenfunktion liegen an festen Orten, im Zeitablauf beobachtet man keine Ausbreitung, sondern nur ein Schwingen. In der *Kundtschen Röhre* oder einer *Glühdrahtröhre* kann man die Knoten bzw. Bäuche stehender Wellen in einer Luftsäule sichtbar machen. An den Schwingungsbäuchen wird feiner Staub in Bewegung versetzt bzw. der Glühdraht stärker gekühlt, so daß sich die Schwingungsknoten hervorheben.

2.3.1.6. Interferenz. Unter Interferenz versteht man die Überlagerung von Wellen. Interferenzerscheinungen können aber nur beobachtet werden, wenn sich die Phasenunterschiede der Teilwellen in der Beobachtungszeit nicht ändern. Zwei Wellen mit dem Phasenunterschied φ verstärken oder schwächen sich je nach dem Wert von φ:
$\cos(kx - \omega t + \varphi/2) + \cos(kx - \omega t - \varphi/2)$
$= 2\cos(\varphi/2)\cos(kx - \omega t)$:
Verstärkung für $|\cos(\varphi/2)| > 1/2$ bzw. $|\varphi| < 120°$,
Schwächung für $|\cos(\varphi/2)| < 1/2$ bzw. $|\varphi| > 120°$,
Auslöschung für $\cos(\varphi/2) = 0$ bzw. $\varphi = \pm 180° = \pm\pi$.
Man kann den Phasenunterschied in eine Wegstrecke Δx umrechnen, $\varphi = k\Delta x$, $\Delta x = \lambda(\varphi/2\pi)$, um die die beiden interferierenden Wellen gegeneinander verschoben sind. Auslöschung entspricht dann einem Wegunterschied $\lambda/2$. Solche Interferenzen können z. B. an dünnen Ölfilmen beobachtet werden, wo sich die an den beiden Grenzflächen, Öl/Luft bzw. Öl/Wasser, reflektierten Wellen überlagern. Je nach dem Wegunterschied tritt für eine Wellenlänge Auslöschung auf. Bei variierender Dicke des Filmes entstehen farbige Figuren.

Interferenzbilder lassen sich leicht mit Wasserwellen demonstrieren. Senden zwei Punktquellen bei $\pm \boldsymbol{a}/2$ Kreiswellen aus $u \sim \mathrm{e}^{-i\omega t}(\mathrm{e}^{ik|r-a/2|} + \mathrm{e}^{ik|r+a/2|}) \approx \mathrm{e}^{i(kr-\omega t)}\, 2\cos(k\boldsymbol{e}\cdot\boldsymbol{a}/2)$, dann gibt es je nach Phasenunterschied Orte $\boldsymbol{r} = e\boldsymbol{r}$ mit Verstärkung und Schwächung (Abb. 2.3.1–7). Für Abstände $r \gg a$ wird deren Lage durch den Faktor $\cos(kea/2)$ bestimmt, der sich aus dem Wegunterschied $\Delta x = \boldsymbol{e}\cdot\boldsymbol{a}$ (Abb. 2.3.1–8) herleitet. Für $\Delta x = 0\,(\boldsymbol{e} \perp \boldsymbol{a})$ wie auch für $\Delta x = \pm\lambda, \pm 2\lambda$, ... verstärken sich beide Wellen, für $\Delta x = \pm\lambda/2, \pm 3\lambda/2, \ldots$ löschen sie sich aus. Damit Minima beobachtet werden können, muß also $a > \lambda/2$ gelten.

2.3.1.7. Kohärenz. Um Interferenzbilder zu erhalten, muß der Phasenunterschied der interferierenden Wellenzüge

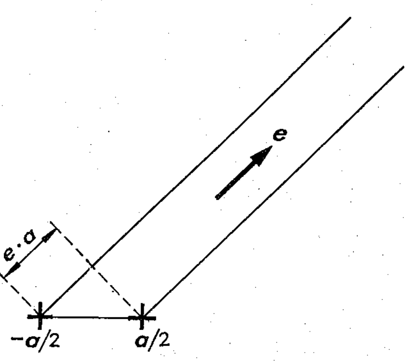


Abb.2.3.1–8 Unterschied des Laufweges. Für einen entfernt liegenden Punkt $\boldsymbol{r}$ kann der Unterschied in der Richtung, in der $\boldsymbol{r}$ von den Zentren bei $\pm\boldsymbol{a}/2$ gesehen wird, vernachlässigt werden

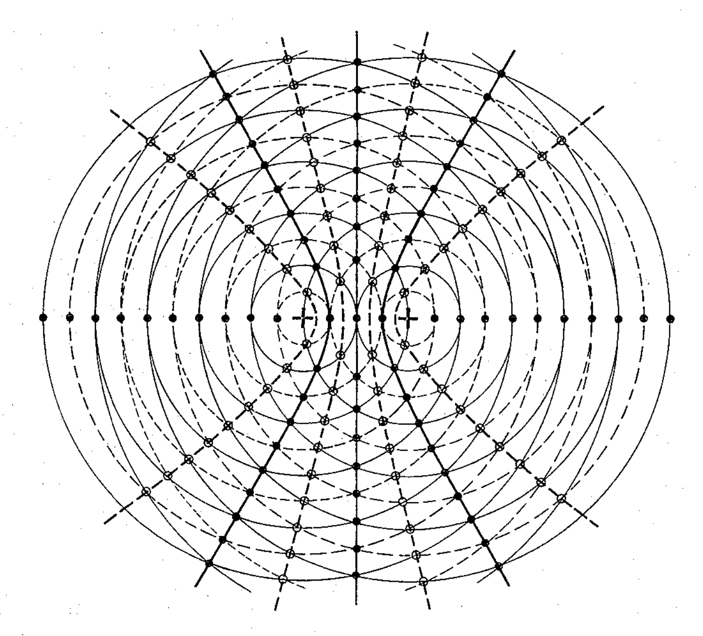


Abb. 2.3.1–7 Wellenberge (———) und -täler (– – –) zweier Kugelwellen, deren Zentren 2λ entfernt liegen. Auslöschung (○) und Verstärkung (●) sind gekennzeichnet

einen festen Wert haben und zeitlich konstant sein. Dies ist jedoch aus verschiedenen Gründen nicht immer erfüllt. Bei der spontanen Emission von Licht, z. B. von einem Glühdraht, sind die Phasenkonstanten der zu einzelnen elementaren Emissionsprozessen gehörenden Wellenzüge regellos verteilt. Die Überlagerung des Lichtes verschiedener Lichtquellen führt daher nicht zu Interferenzbildern.
Der einfachste Weg, zwei interferenzfähige Wellenzüge herzustellen, ist die Teilung einer Welle, z. B. an einem halbdurchlässigen Spiegel. Allerdings darf der Unterschied im Laufweg der beiden Wellenzüge bis zur Überlagerung nicht zu groß sein. Diese *Kohärenzlänge* wird durch die Länge der Wellenzüge bestimmt, die zu einem elementaren Emissionsprozeß gehören. Sind die Quellen, die die einzelnen Wellenzüge aussenden, zu ausgedehnt, dann variiert der Laufwegunterschied, je nachdem, von welchem Punkt der Quelle man ausgeht. Auch dadurch kann das Interferenzbild verwischt oder zerstört werden.
Schwingen alle elementaren Wellenzüge eines Lichtbündels in Phase zueinander, dann spricht man von *kohärentem Licht*. Im *Laser* erfolgt die induzierte Emission im Takt des schon vorhandenen Feldes. Die Laserstrahlung ist kohärent. Das Licht eines Glühdrahtes ist in diesem Sinne inkohärent.
2.3.1.8. Doppler-Effekt. Bewegen sich Quelle und Beobachter relativ zueinander, dann wird eine veränderte Frequenz gemessen. So wird der Klang eines Rennwagenmotors beim Entfernen tiefer empfunden als bei der Annäherung des Wagens.
Bei der Berechnung der Frequenzänderung muß man unterscheiden, ob die Ausbreitung der Welle – wie beim Schall – an ein Medium gebunden ist oder – wie beim Licht – nicht. Sind v_B bzw. v_Q die Geschwindigkeiten, mit denen sich Beobachter bzw. Quelle (Abb. 2.3.1–9) bezüglich des Mediums bewegen ($v > 0$ bei Annäherung), dann gilt für Schallwellen (Schallgeschwindigkeit c) $f = f_0(1 + v_B/c)/(1 - v_Q/c)$. Aus der Entwicklung $1/(1 - v/c) = 1 + (v/c) + (v/c)^2 + \ldots$ erkennt man, daß für einen bewegten Beobachter bzw. eine bewegte Quelle eine in zweiter Ordnung in v/c unterschiedliche Frequenzänderung auftritt.
Gibt es kein Medium, an das die Welle gebunden ist, dann kann nur die Relativgeschwindigkeit v zwischen Beobachter und Quelle angegeben werden, eine Unterscheidung zwischen v_B und v_Q ist unnötig. Aus der Konstanz der Vakuumlichtgeschwindigkeit c in allen Bezugssystemen (vgl. spezielle Relativitätstheorie) folgt für Lichtwellen $f = f_0\sqrt{(1 + v/c)/(1 - v/c)}$ ($v > 0$ bei Annäherung). Die Unterschiede zwischen beiden Formeln machen sich erst in zweiter Ordnung in v/c bemerkbar.
Bewegt sich eine Schallquelle mit einer Geschwindigkeit, die über der Schallgeschwindigkeit liegt, dann eilt der Quelle kein Schallfeld voraus. Das Schallfeld liegt innerhalb des *Machschen Kegels* (Abb. 2.3.1–10). Der halbe Öffnungswinkel dieses Kegels ist durch das Verhältnis $v/c = \sin\alpha$ bestimmt. Erreicht der Kegelmantel den Ort eines Beobachters, dann vernimmt dieser einen scharfen Knall.

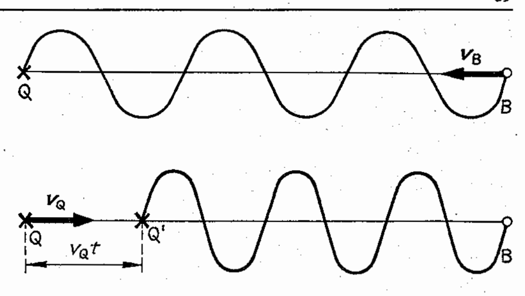


Abb. 2.3.1–9 Doppler-Effekt

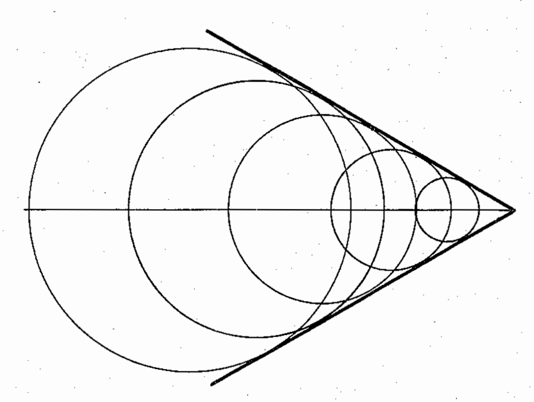


Abb. 2.3.1–10 Machscher Kegel

2.3.2. Wellengleichungen

Die räumliche Verteilung und der zeitliche Ablauf von Feldern wird durch Grundgleichungen (Wellengleichungen) bestimmt, die mathematisch gesehen partielle Differentialgleichungen sind. Sie enthalten Orts- und Zeitableitungen der *Wellenfunktion*. Diese Beschreibung entspricht der Vorstellung der *Nahwirkung*. Änderungen eines Feldes breiten sich mit endlicher Geschwindigkeit aus. In einer kleinen Zeitspanne $\mathrm{d}t$ wirkt sich der Wert des Feldes in einem Punkt $\boldsymbol{r}$ nur auf dessen infinitesimale Umgebung aus. Die Verwendung von Kraftgesetzen $\boldsymbol{F}(\boldsymbol{r}_1, \boldsymbol{r}_2)$, wie für die Gravitationskraft, die nur die Orte $\boldsymbol{r}_1$, $\boldsymbol{r}_2$ der beiden Massen enthalten, entspricht der Vorstellung der *Fernwirkung*. Eine Änderung der Lage $\boldsymbol{r}_2$ macht sich sofort in einer Änderung des Wertes des Kraftfeldes am Ort $\boldsymbol{r}_1$ bemerkbar. Die Fernwirkungsvorstellung enthält also implizit die Annahme einer unendlich großen Ausbreitungsgeschwindigkeit.
2.3.2.1. Superponierbarkeit. Die Wellenfunktionen verschiedener Wellenzüge lassen sich beim Durchdringen überwiegend einfach addieren, d. h., ist jeder einzelne Wellenzug eine Lösung der Wellengleichung, dann ist auch die Summe der beiden Wellenfunktionen eine Lösung der Wellengleichung. Die Lösungen lassen sich superponieren. Daraus folgt, daß die zugehörige Wellengleichung linear ist, d. h. die Wellenfunktion nur in erster Ordnung enthält. Nichtlineare Effekte vgl. 2.3.6.
2.3.2.2. Wellenausbreitung in homogenen Medien. In einem räumlich homogenen und isotropen Medium ist kein Raumpunkt und keine Raumrichtung ausgezeichnet. Ohne zeitabhängige Eingriffe ist ebenfalls kein Zeitpunkt ausgezeichnet. Daher kann in diesem Falle die Wellengleichung die Koordinaten $\boldsymbol{r}$ und t nicht explizit enthalten, es können nur die Ableitungen $\partial/\partial t$ bzw. $\partial/\partial \boldsymbol{r} = \boldsymbol{e}_x\partial/\partial x + \boldsymbol{e}_y\partial/\partial y + \boldsymbol{e}_z\partial/\partial z$ auftreten.

$$L(\partial/\partial t, \partial/\partial \boldsymbol{r})\, u(\boldsymbol{r}, t) = 0$$

Diese *lineare Wellengleichung* hat die ebene Welle $u(\boldsymbol{r}, t)$

$= u_0 \mathrm{e}^{-\mathrm{i}(\omega t - kr)}$ als spezielle Lösung, falls zwischen Kreisfrequenz und Wellenvektor der durch die Wellengleichung, nämlich durch $L(-\mathrm{i}\omega, \mathrm{i}k) = 0$, bestimmte Zusammenhang $\omega = \omega(k)$ besteht.

Über das betrachtete physikalische Phänomen wurde hierbei keinerlei Annahme gemacht, diese Aussage gilt generell. Es wird daraus verständlich, daß ebene Wellen eine weitverbreitete Form der Wellenausbreitung sind. Verschiedene physikalische Erscheinungen unterscheiden sich in der Bedeutung der Wellenfunktion und durch den Zusammenhang zwischen Kreisfrequenz und Wellenvektor (oder Frequenz und Wellenlänge).

2.3.2.3. Formen von Wellengleichungen. Die Wellengleichung für *Seilwellen* lautet $[(1/c^2)(\partial^2/\partial t^2) - (\partial^2/\partial x^2)]\, \boldsymbol{u}(x, t) = 0$, $\omega = ck, \lambda f = c$. c ist die Ausbreitungsgeschwindigkeit der Seilwellen und $\boldsymbol{u}$ bedeutet den Vektor der Auslenkung, der senkrecht zur Ausbreitungsrichtung (x-Achse) liegt.

Bei *Schallwellen* in Gasen und Flüssigkeiten mit $[(1/c^2)(\partial^2/\partial t^2) - \Delta]\, u(\boldsymbol{r}, t) = 0$, $\omega = c|\boldsymbol{k}|, \lambda f = c$ bedeutet c die Schallgeschwindigkeit. Aus der skalaren Wellenfunktion u lassen sich Druckschwankungen $p - \bar{p} = \varrho(\partial u/\partial t)$ und Geschwindigkeit (Schallschnelle) $\boldsymbol{v} = -\partial u/\partial \boldsymbol{r}$ berechnen. In Festkörpern (vgl. Gitterschwingungen) gibt die Wellenfunktion die Auslenkungen an. Zu einem Wellenvektor gibt es drei Polarisationsrichtungen, im einfachsten Fall eine longitudinale und zwei transversale Wellen.

Für *elektromagnetische Wellen* tritt in der Wellengleichung $[n^2/c^2(\partial^2/\partial t^2) - \Delta]\, u(\boldsymbol{r}, t) = 0$, $\omega = (c/n)|\boldsymbol{k}|, \lambda f = c/n$, die Vakuumlichtgeschwindigkeit c auf bzw. bei Ausbreitung im Medium mit dem Brechungsindex n die Geschwindigkeit c/n. u steht stellvertretend für das skalare Potential U oder das Vektorpotential $\boldsymbol{A}$, aus denen sich die Felder $\boldsymbol{E} = -\partial U/\partial \boldsymbol{r} - \partial \boldsymbol{A}/\partial t$ bzw. $\boldsymbol{B} = \partial/\partial \boldsymbol{r} \times \boldsymbol{A}$ (= rot $\boldsymbol{A}$) gewinnen lassen.

Elektronenwellen genügen der Wellengleichung $[-(\hbar/\mathrm{i})(\partial/\partial t) - \hbar^2/(2m)\Delta]\, u(r, t) = 0$, $\hbar\omega = (\hbar k)^2/(2m)$. u hat hier die Bedeutung einer Wahrscheinlichkeitsamplitude, die der Normierungsbedingung $\int \mathrm{d}\tau |u^2| = 1$ genügt.

In die Wellengleichung sind gegebenenfalls *Quellterme* einzubauen. Sie sind in den Raumbereichen von Null verschieden, in denen Wellen erzeugt werden, bei elektromagnetischen Wellen z. B. an den Orten der Ladungen und Ströme.

Das betrachtete physikalische Gebiet ist berandet. Um eine eindeutige Lösung der Wellengleichung zu gewinnen, müssen daher außer der Anfangsverteilung *Randbedingungen* zu allen Zeiten angegeben werden. Diese bestehen i. allg. darin, daß man die Wellenfunktion oder deren Normalableitung (oder eine kombinierte Aussage) auf dem Rande vorgeben muß.

Ist das betrachtete System stückweise homogen, dann kann man in jedem Teilsystem die Lösungen des jeweiligen homogenen Systems betrachten. In ihnen treten dann die Ausbreitungsgeschwindigkeit (oder andere Materialkonstanten) des Teilsystems auf. Die Wellengleichung bestimmt die *Grenzbedingungen*, durch die die Teillösungen miteinander angeschlossen werden müssen.

2.3.2.4. Bilanzgleichungen, Erhaltungsgrößen. Der Wert einer extensiven Größe W, z. B. der inneren Energie, in einem Volumen V kann sich im Zeitablauf dadurch ändern, daß 1. ein Strom J (Energiestrom) durch die Oberfläche des Volumens austritt und 2. im Volumen ein Verlust N (Verlustleistung) zu verzeichnen ist.

Die *Bilanzgleichung* lautet daher

$\mathrm{d}W/\mathrm{d}t + J + N = 0$ oder $\partial w/\partial t + \mathrm{div}\boldsymbol{j} + \nu = 0$.

Die auf das Volumenelement bezogene differentielle Form ergibt sich, wenn man die Dichte w, $W = \int \mathrm{d}\tau\, w$, Stromdichte $\boldsymbol{j}$, $J = \oint \mathrm{d}\boldsymbol{a} \cdot \boldsymbol{j}$, und die Verlustdichte ν, $N = \int \mathrm{d}\tau \nu$, einführt. Betrachtet man die *Ladung*, dann ist N bzw. ν gleich Null, da Ladung nicht erzeugt oder vernichtet werden kann. J bedeutet die elektrische Stromstärke, $\boldsymbol{j}_e$ die elektrische Stromdichte. In der Bilanz für die *elektromagnetische Energie* ist $\boldsymbol{j}$ der Poynting-Vektor. Die Verlustleistungsdichte $\nu = \boldsymbol{j}_e \cdot \boldsymbol{E}$ beschreibt die Umwandlung von elektromagnetischer Energie in mechanische oder Wärmeenergie. Tritt kein Strom aus dem Volumen heraus und kein Verlust auf, dann ist die betrachtete physikalische Größe eine *Erhaltungsgröße*, der im betrachteten Volumen vorhandene Wert ändert sich im Zeitablauf nicht.

Bilanzgleichungen, d.h. die Zusammenhänge von W, J und N mit der Wellenfunktion, müssen aus der Wellengleichung gewonnen werden. So folgt aus der Wellengleichung für den Schall eine Bilanzgleichung mit

$$W = \int \mathrm{d}\tau \frac{\varrho}{2}\left[\left(\frac{1}{c}\,\frac{\partial u}{\partial t}\right)^2 + \left(\frac{\partial u}{\partial \boldsymbol{r}}\right)^2\right],$$

$$J = -\oint \mathrm{d}\boldsymbol{a}\,\varrho \frac{\partial u}{\partial t}\,\frac{\partial u}{\partial \boldsymbol{r}}$$

als Ausdruck für Energie bzw. Energiestrom (ϱ = Massendichte). Der von einem Wellenpaket transportierte Impuls ist betragsmäßig $P = W/c$. Dieser Zusammenhang gilt auch für Lichtwellen, die ja formal derselben Wellengleichung genügen. Da jedoch die Lichtgeschwindigkeit wesentlich größer ist als die Schallgeschwindigkeit, ist bei gleichem Energiestrom der Strahlungsdruck bei Schallwellen wesentlich größer.

2.3.2.5. Retardierte Lösungen. Eine Quellverteilung $q(\boldsymbol{r}, t)$ stellt eine Inhomogenität der Wellengleichung dar $\left[\frac{1}{c^2}\,\frac{\partial^2}{\partial t^2} - \Delta\right] u(\boldsymbol{r}, t) = q(\boldsymbol{r}, t)$. Die Lösung dieser Gleichung, die im Unendlichen abklingt, kann angegeben werden.

Für eine statische Punktquelle der Quellstärke Q im Volumenelement bei $\boldsymbol{r} = 0\,(q(\boldsymbol{r}, t) = Q\delta(\boldsymbol{r}))$ ist sie $u(\boldsymbol{r}) = Q/(4\pi r)$. Wegen der Linearität der Wellengleichung kann man daraus die Lösung für eine beliebige zeitunabhängige Quellverteilung $q(\boldsymbol{r})$ aufbauen

$$u(\boldsymbol{r}) = \int \mathrm{d}\tau' \frac{q(\boldsymbol{r}')}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}.$$

Ist die Quellstärke zeitabhängig, dann enthält auch die Wellenfunktion die Zeit explizit

$$u(\boldsymbol{r}, t) = \int \mathrm{d}\tau' \frac{q(\boldsymbol{r}', t - |\boldsymbol{r} - \boldsymbol{r}'|/c)}{4\pi|\boldsymbol{r} - \boldsymbol{r}'|}.$$

Die Stärke des Feldes zur Zeit t am Aufpunkt $\boldsymbol{r}$ wird aber durch die Quellstärke zur Zeit $t - |\boldsymbol{r} - \boldsymbol{r}'|/c$ bestimmt. Der

Abb. 2.3.2-1 Entwicklung von $|\boldsymbol{r} - \boldsymbol{r}'|$

Zusatz $|r - r'|/c$, die *Retardierung*, bedeutet die Laufzeit von der Quelle am Punkt r' bis zum Aufpunkt r bei der Ausbreitungsgeschwindigkeit c. In großer Entfernung vom Quellgebiet kann $|r - r'|$ durch $r - e \cdot r'$ (Abb. 2.3.2–1) angenähert werden. Im Nenner braucht nur die nullte Näherung $|r - r'| \approx r$ berücksichtigt zu werden, da nur die zu $1/r$ proportionalen Feldanteile einen endlichen Beitrag zur Energiestromdichte liefern (vgl. 2.1.3.4.). In der Retardierung beschreibt der Term er'/c den Unterschied in der Laufzeit der Wellen, die von den verschiedenen Punkten r' der Quellverteilung ausgehen. Damit bestimmt er die Interferenzeffekte. Die Interferenzeffekte sind schwach, wenn $|\omega er'/c|$ klein gegen 2π ist. Das ist dann der Fall, wenn die Ausdehnung der Quellverteilung klein gegen die Wellenlänge ist oder wenn die Geschwindigkeit der Teilchen in der Quelle klein gegen die Ausbreitungsgeschwindigkeit ist.

2.3.3. Wellenpakete

Eine ebene Welle ist ein idealisierter Grenzfall einer Welle, sie ist über den gesamten Raum ausgebreitet und hat eine orts- und zeitunabhängige Amplitude. Bei einer lokalisierten Feldverteilung mit einer charakteristischen Ausdehnung Δx spricht man von einem Wellenpaket.

2.3.3.1. Formunveränderte Ausbreitung. Wellengleichungen vom Typ $[(1/c^2)(\partial^2/\partial t^2) - \Delta]u(x, t) = 0$, wie sie bei Seil-, Schall- und Lichtwellen auftreten, haben als mögliche Lösung $u(x, t) = g(x - ct)$. Das bedeutet, daß die durch $g(x)$ bestimmte Form des Wellenpaketes im Zeitablauf erhalten bleibt (Abb. 2.3.3–1). Es bewegt sich mit der Geschwindigkeit c im betrachteten Fall nach rechts), d.h., die *Gruppengeschwindigkeit* ist c. Abweichungen von der formunveränderten Ausbreitung haben als Ursachen 1. die *Dispersion* (vgl. 2.3.3.4.) und 2. die *Dämpfung* (vgl. 2.3.3.5.).

2.3.3.2. Gruppen und Phasengeschwindigkeit. Überlagert man zwei ebene Wellen $u = u_0 \cos(kx - \omega t)$ zu zwei benachbarten Wellenzahlen $k = k_0 \pm \Delta k$, dann erhält man für $t = 0$ eine *amplitudenmodulierte* ebene Welle $u(x, 0) = 2u_0 \cos \Delta kx \cos k_0 x$ (Abb. 2.3.3–2). Will man den Zeitablauf verfolgen, dann muß man beachten, daß in den beiden Anteilen unterschiedliche Frequenzen $\omega(k_0 \pm \Delta k) = \omega_0 \pm \Delta k(\mathrm{d}\omega/\mathrm{d}k)_0$ auftreten. Das führt auf
$u(x, t) = 2u_0 \cos \Delta k(x - v_{gr}t) \cos k_0(x - v_{ph}t)$, $v_{gr} = (\mathrm{d}\omega/\mathrm{d}k)_0$, $v_{ph} = \omega_0/k_0$, worin die Phasengeschwindigkeit v_{ph} und die Gruppengeschwindigkeit die angegebene Bedeutung haben. v_{gr} bestimmt, mit welcher Geschwindigkeit sich die Amplitudenmodulation bewegt, v_{ph} gibt die Phasengeschwindigkeit der zur Wellenzahl k_0 gehörenden ebenen Welle an. Für $\omega = ck$ gilt $v_{gr} = v_{ph}$. Darüber hinaus sind beide unabhängig von k. Für Lichtwellen in Materie gilt $\omega = ck/n$ mit dem frequenzabhängigen Brechungsindex $n(\omega)$. Diese Abhängigkeit ist so beschaffen, daß stets $v_{gr} \leqq c$ (Vakuumlichtgeschwindigkeit) gilt. v_{ph} kann größer als c werden.

2.3.3.3. Fourier-Zerlegung. Unter einer Fourier-Zerlegung (vgl. 13.1.4.1.) versteht man eine Entwicklung nach Sinus- und Cosinusfunktionen, deren Argumente Grund- und Oberschwingungen charakterisieren. Eine im Intervall $0 \leqq x \leqq L$ mit $g(0) = g(L) = 0$ gegebene Funktion (z.B. Auslenkung einer Saite) ist beschreibbar durch

$$g(x) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi}{L} x = \sum_{n=1}^{\infty} b_n \sin k_n x.$$

Die Grundschwingung gehört zur Wellenlänge $2L$. Daneben treten *Oberschwingungen* zu Wellenlängen $2L/n$ auf.

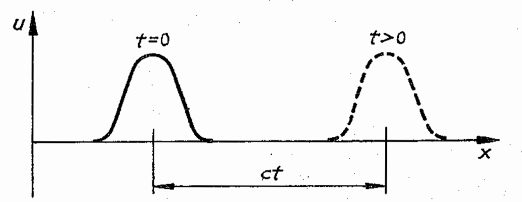


Abb. 2.3.3–1 Formunveränderte Ausbreitung

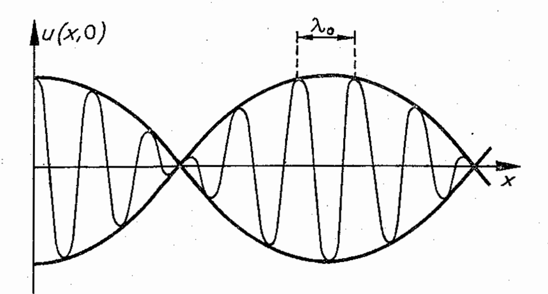


Abb. 2.3.3–2 Amplitudenmodulation bei der Überlagerung zweier Wellen $\lambda = \lambda_0 \pm \Delta\lambda$

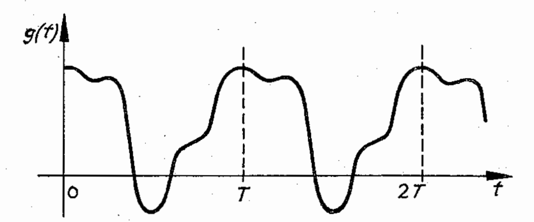


Abb. 2.3.3–3 Mit Intervall T zeitlich periodische Funktion

Eine im Intervall T zeitlich periodische Funktion ist in

$$g(t) = a_0 + \sum_{n=1}^{\infty} (a_n \cos n\omega_0 t + b_n \sin n\omega_0 t),$$

$\omega_0 = 2\pi/T$ zerlegbar (Abb. 2.3.3–3). Es treten ganzzahlige Vielfache $\omega = n\omega_0$ der zur Schwingungsdauer T gehörigen Frequenz auf. Geht die Intervallänge gegen unendlich, dann treten bei der Fourier-Entwicklung alle Wellenlängen auf. Man spricht von einem *Fourier-Integral*. Am zweckmäßigsten ist die komplexe Darstellung

$$g(x) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} \tilde{g}(k)\, \mathrm{e}^{\mathrm{i}kx}, \quad \tilde{g}(k) = \int_{-\infty}^{+\infty} \mathrm{d}x\, g(x)\, \mathrm{e}^{-\mathrm{i}kx}.$$

$\tilde{g}(k)$ ist die *Fourier-Transformierte* zu $g(x)$. Ist $g(x)$ reell, dann gilt $\tilde{g}^*(-k) = g(k)$.

Eine beliebige Anfangswellenfunktion $u(x, 0) = g(x)$ kann man als Fourier-Integral schreiben. Die zeitabhängige Lösung ist dann

$$u(x, t) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} \tilde{g}(k)\, \mathrm{e}^{\mathrm{i}(kx - \omega(k)t)}.$$

Ein beliebiges Wellenpaket ist eine Überlagerung von ebenen Wellen aller Wellenzahlen, wobei die positiven nach rechts und die negativen nach links laufende Wellen be-

schreiben. Die Variation des Betrages von k erfaßt alle Wellenlängen.

2.3.3.4. Unschärferelation. Bei der Amplitudenmodulation aus Abb. 2.3.3–2 treten Knoten im Abstand $\pi/\Delta k$ auf, die Welle bleibt aber noch über die gesamte x-Achse ausgebreitet. Will man ein Wellenpaket (Abb. 2.3.3–4) aufbauen, das auf eine Breite Δx lokalisiert ist, dann muß man ebene Wellen aus einem Wellenzahlbereich überlagern. Setzt man in die Fourier-Zerlegung der Wellenfunktion die Spektralfunktion $\tilde{g}(k) = \sqrt{\frac{2\pi}{\Delta k}}\, e^{-ikx_0} \exp\left[-\frac{1}{2}\left(\frac{k-k_0}{\Delta k}\right)^2\right]$ ein, dann erhält das Wellenpaket die Form

$$u(x, 0) = \frac{1}{\sqrt{2\pi}\Delta x}\, e^{ik_0(x-x_0)} \exp\left[-\frac{1}{2}\left(\frac{x-x_0}{\Delta x}\right)^2\right].$$

Die Spektralfunktion der Breite Δk im Wellenzahlraum führt auf ein Wellenpaket der Breite $\Delta x = 1/\Delta k$ im Ortsraum. Man kann leicht Spektralfunktionen größerer Ausdehnung angeben, die zu keiner Verschmälerung des Wellenpaketes führen, aber jede Verengung der betrachteten Spektralfunktion führt zu einem im Ortsraum breiteren Wellenpaket. Für das Produkt aus *Breite der Spektralfunktion im Wellenzahlraum* und *Breite des Wellenpaketes im Ortsraum* gibt es eine untere Schranke $\Delta x \Delta k \geqq 1$. Der auftretende Zahlenwert hängt von der Definition von Δx und Δk ab, hier wurden sie als Breite der Gauß-Funktion eingeführt. Wesentlich ist, daß eine *untere Schranke* für das Produkt existiert.

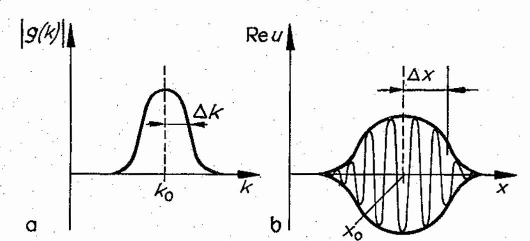


Abb. 2.3.3–4 (a) Spektralfunktion und (b) Verteilung im Ortsraum

Man kann durch Auswertung des Fourier-Integrals auch den *Zeitablauf des Wellenpaketes* bestimmen

$$u(x, t) = \frac{1}{\sqrt{\widetilde{\Delta x}\, 2\pi}}\, e^{ik_0(x - x_0 - v_{ph}t)} \exp\left[-\frac{1}{2}\left(\frac{x - x_0 - v_{gr}t}{\widetilde{\Delta x}}\right)^2\right].$$

Gruppen- und Phasengeschwindigkeit sind wie in 2.3.3.2. definiert. Die Breite des Wellenpaketes $\Delta x(t)$ ist i. allg. zeitabhängig, es läuft auseinander. $\Delta x(t)$ enthält $d^2\omega/dk^2$. $d^2\omega/dk^2 \neq 0$ bedeutet aber auch, daß die Gruppengeschwindigkeit von der Frequenz oder Wellenzahl abhängt, d. h., daß *Dispersion* auftritt. Das Auseinanderlaufen ist damit unmittelbar verknüpft.

2.3.3.5. Dämpfung. Wellen und Schwingungen klingen zeitlich oder/und räumlich ab, da die von der Welle transportierte Energie durch Reibung oder Abstrahlung aufgezehrt wird (*Dissipation*). Dabei wird überwiegend ein exponentielles Abklingen beobachtet.

So wird der Zeitablauf einer schwingenden Saite durch $u(x, t) = u_0 e^{-\gamma t/2} \cos \omega t \sin(\pi x/L)$ dargestellt. Innerhalb einer Zeit $\tau = 1/\gamma$ ist die Schwingungsenergie ($\sim u^2$) auf den e-ten Teil abgeklungen.

Das räumliche Abklingen einer Welle, die in eine Substanz eindringt, wird durch $u(x, t) = u_0 e^{-\varkappa x} \cos(kx - \omega t)$ beschrieben. Die Amplitude klingt exponentiell ab. In der komplexen Darstellung kann man diese Abklingfaktoren durch Imaginärteile der Wellenzahl $k \to k + i\varkappa$ oder der Kreisfrequenz $\omega \to \omega - i\gamma/2$ erfassen.

Wird die Dämpfung in die Wellengleichung einbezogen, dann kann sie für periodische Wellen durch einen *komplexen Brechungsindex* $N(\omega)$ berücksichtigt werden, so daß die Wellengleichung $[N^2(\omega)/c^2 \partial^2/\partial t^2 - \Delta]u(\boldsymbol{r}, t) = 0$ (für Licht- oder Schallwellen) lautet. Der Realteil von $N = n + ia$ wird als *Brechungsindex*, der Imaginärteil als *Absorptionskoeffizient* bezeichnet.

Die Ausbreitung von Wellen in einem homogenen Medium wird also durch zwei Funktionen der Frequenz charakterisiert. Statt n und a verwendet man – insbesondere bei elektromagnetischen Wellen – auch Real- und Imaginärteil der *dielektrischen Funktion*, die über $N^2 = \varepsilon = \varepsilon_1 + i\varepsilon_2$ mit dem komplexen Brechungsindex zusammenhängt.

2.3.3.6. Kramers-Kronig-Relationen. Wird die Reaktion einer Substanz auf ein periodisches Feld untersucht, z. B. das Auftreten einer Polarisation $P = \varepsilon_0 \chi E$ in einem elektrischen Feld E, dann treten zwei Funktionen der Frequenz auf, z. B. Real- und Imaginärteil der *Suszeptibilität* $\chi = \chi_1 + i\chi_2$ oder der dielektrischen Funktion $\varepsilon = 1 + \chi$. Sie beschreiben die Änderung der Ausbreitungsgeschwindigkeit durch das Mitschwingen der Ladungen und die Dämpfung durch Umwandlung von elektromagnetischer Energie in Wärmeenergie bzw. eine andere Energieform.

Zwischen diesen beiden Funktionen besteht ein Zusammenhang, der allein aus der *Kausalitätsforderung*, d. h. aus der Forderung, daß eine Reaktion nicht vor dem verursachenden Feld auftreten kann, abgeleitet werden kann. Dieser Zusammenhang lautet

$$\chi_1(\omega) = \frac{2}{\pi} \int_0^\infty d\Omega\, \Omega \frac{\chi_2(\Omega)}{\Omega^2 - \omega^2},$$

$$\chi_2(\omega) = -\frac{2\omega}{\pi} \int_0^\infty d\Omega \frac{\chi_1(\Omega)}{\Omega^2 - \omega^2},$$

wobei der Hauptwert des Integrals zu nehmen ist. Zusammenhänge dieser Art werden als Kramers-Kronig-Relationen bezeichnet.

2.3.4. Wellenausbreitung in inhomogenen Medien

In inhomogenen Medien sind die Materialkonstanten (hier Brechungsindex und Absorptionskoeffizient) Funktionen des Ortes. Der einfachste, aber auch wichtigste Fall ist der eines stückweise homogenen Mediums. Die Inhomogenität kommt dann darin zum Ausdruck, daß die Materialkonstanten an den Grenzflächen zwischen den verschiedenen Materialien ihren Wert sprunghaft ändern. Trifft eine Welle auf eine *Grenzfläche*, so interessiert zunächst der reflektierte und durchgelassene Anteil. Wegen der endlichen Ausdehnung der Grenzfläche, d.h. dadurch, daß Ränder vorhanden sind, treten Beugungseffekte auf. Von der Absorption wird bei den folgenden Betrachtungen in der Regel abgesehen.

2.3.4.1. Reflexion und Brechung. Für *Seilwellen* sollen nur die beiden Grenzfälle besprochen werden, in denen das Ende fest eingespannt ist oder frei hängt. Wird die x-Achse so gewählt, daß das Seil im Bereich $x < 0$ liegt (Abb. 2.3.4–1), dann ist die Auslenkung bei formunveränderter Ausbreitung, d.h. bei $\lambda f = c$ bzw. $\omega = ck$, $u(x, t) = g(ct - x) + Rg(ct + x)$. Bei eingespanntem Ende ist $u(0, t) = 0$, d. h.

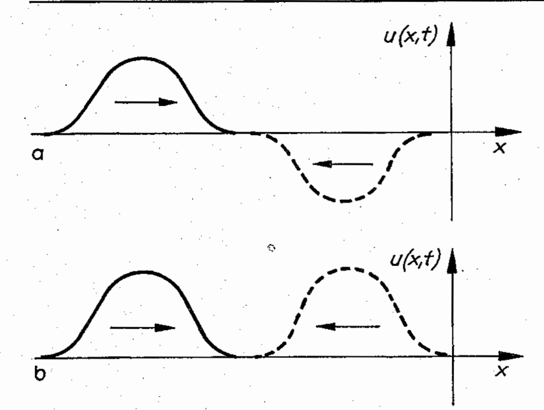


Abb. 2.3.4–1 Reflexion am (a) festen und (b) losen Ende

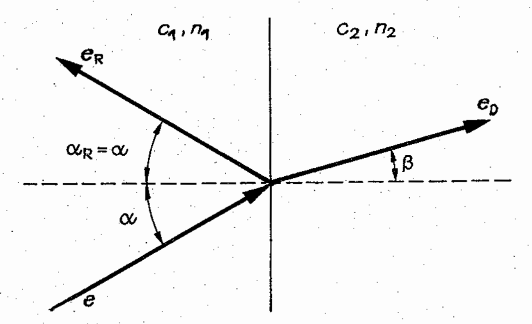


Abb. 2.3.4–2 Reflexion und Brechung für $n_2 > n_1$

der *Reflexionskoeffizient* $R = -1$. Bei losen Ende befindet sich bei $x = 0$ ein Schwingungsbauch, der Reflexionskoeffizient hat den Wert $R = 1$.

Bei der Wellenausbreitung im Dreidimensionalen interessiert nicht nur der durch R bzw. T charakterisierte reflektierte bzw. durchgelassene Anteil der Welle, sondern auch die Ausbreitungsrichtung dieser Anteile, d. h. der *Reflexionswinkel* α_R bzw. der *Brechungswinkel* β (Abb. 2.3.4–2). Die Ebene, in der die Normalenrichtung der Grenzfläche und die Ausbreitungsrichtung $\boldsymbol{k} = k\boldsymbol{e}$ der einfallenden ebenen Welle liegen, wird als *Einfallsebene* bezeichnet.

An einer Grenzfläche muß die Wellenfunktion immer bestimmte Anschlußbedingungen erfüllen. So müssen bei der Schallwelle Druck und Normalkomponente der Schallschnelle stetig verlaufen. Allein aus der Existenz solcher Bedingungen folgt, daß 1. die Ausfalls- und Einfallsebene übereinstimmen, 2. der Reflexionswinkel gleich dem Einfallswinkel ($\alpha_R = \alpha$) ist und 3. der Brechungswinkel sich aus dem *Snelliusschen Brechungsgesetz* ergibt,

$$n_2 \sin\beta = n_1 \sin\alpha \quad \text{bzw.} \quad \sin\beta = (c_2/c_1)\sin\alpha.$$

Diese Aussagen sind für alle Wellenausbreitungen gleich. Beim Übergang zum „dichteren" Material ($n_2 > n_1$) wird die Welle zum Lot hin gebrochen, es gilt $\beta < \alpha$. Beim Übergang zum „dünneren" Material wird die Welle vom Lot weg gebrochen, es gilt $\beta > \alpha$. In diesem Fall gibt es für den Einfallswinkel einen Grenzwert $\sin\alpha = n_2/n_1$, für den der Brechungswinkel den Wert 90° annimmt. Für flacheren Einfall, d. h. größere Einfallswinkel, tritt *Totalreflexion* auf.

Die Ausdrücke für den Reflexions- und Durchlaßkoeffizienten hängen von der Form der Anschlußbedingungen und damit von der Art der Wellenausbreitung ab. Für das Druckfeld p (vgl. 2.3.2.3.) von Schallwellen ergibt sich bei senkrechtem Einfall $R = (z_2 - z_1)/(z_2 + z_1)$, $T = (2z_2)/(z_2 + z_1)$ ($z = \varrho c$ = Wellenwiderstand, ϱ = Massendichte). Der Schall wird an einer Grenzfläche nicht reflektiert, wenn die Wellenwiderstände der aneinanderstoßenden Medien gleich groß sind. Sind die Wellenwiderstände stark verschieden, dann tritt fast Totalreflexion auf, bei $z_2 \gg z_1$ wird $R \approx 1$, bei $z_2 \ll z_1$ wird $R \approx -1$. Das entspricht den beiden bei der Seilwelle – festes und loses Ende – angeführten Fällen.

Für beliebige Einfallswinkel ergibt sich

$$R = (z_2 \cos\alpha - z_1 \cos\beta)/(z_2 \cos\alpha + z_1 \cos\beta),$$
$$T = (2z_2 \cos\alpha)/(z_2 \cos\alpha + z_1 \cos\beta)$$

für Reflexions- und Transmissionskoeffizienten. Im Bereich der Totalreflexion ist $|R| = 1$.

Die angeführten Formeln gelten für den Fall, daß das Medium 2 einen Halbraum ausfüllt. Das Reflexionsvermögen eines beliebigen Systems, z. B. einer Schicht endlicher Dicke oder eine Folge von Schichten verschiedener Materialien, charakterisiert man durch seine *Impedanz* z. Diese ist so definiert, daß man den Reflexionskoeffizienten (für senkrechten Einfall) in der Form $R = (z - z_1)/(z + z_1)$ schreibt, d.h., ein Halbraum eines Materiales mit dem Wellenwiderstand z hat das gleiche Reflexionsvermögen wie das betrachtete System.

Für Lichtwellen sind Reflexions- und Durchlaßkoeffizient von der Polarisationsrichtung abhängig. Sie sind durch die Fresnelschen Formeln gegeben.

2.3.4.2. Beugung. Trifft eine Welle auf eine Wand mit einem Loch oder stellt man ein Hindernis in die Welle, beobachtet man einen von dem Loch ausgehenden Strahl oder einen Schatten hinter dem Hindernis. Man spricht von der *geradlinigen Ausbreitung* der Strahlen. Dies trifft jedoch nur zu, wenn die Wellenlänge klein gegen die charakteristischen Abmessungen ist. Im allgemeinen müssen Beugungs-

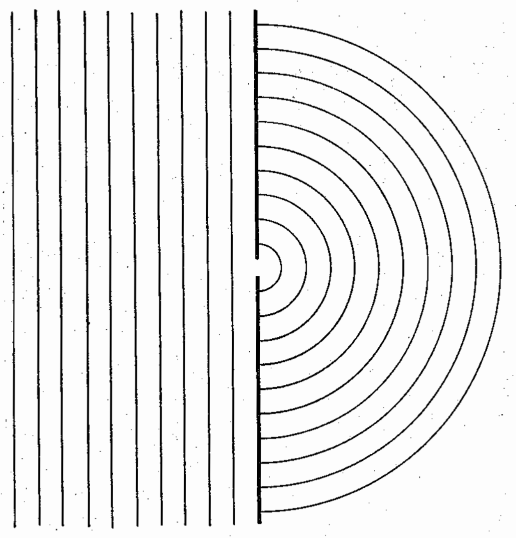


Abb. 2.3.4–3 Eine ebene Welle trifft auf eine Wand mit Loch, und es entsteht eine Kugelwelle

erscheinungen berücksichtigt werden. Beim Auftreten von Interferenzbildern kann man diese besonders gut beobachten.

Die Beugungserscheinungen kann man ausgehend vom *Huygensschen Prinzip* verstehen. Es besagt, daß die Ausbreitung einer Wellenfront erklärt werden kann, indem man jeden Punkt dieser Wellenfront als Ausgangspunkt einer Kugelwelle ansieht. Die einfachste diesem Prinzip entsprechende Beobachtung kann man vornehmen, wenn man eine Welle auf eine Wand fallen läßt, in der sich ein im Vergleich zur Wellenlänge kleines Loch befindet. Auf der anderen Seite der Wand beobachtet man dann eine von diesem Loch ausgehende Kugelwelle (Abb. 2.3.4–3).

Beugung am Spalt (Abb. 2.3.4–4). Hierbei sind alle Punkte des Spaltes als Ausgangspunkt von Kugelwellen anzusehen. Laufen diese Kugelwellen phasengleich aus, was bei senkrechtem Einfall gegeben ist, dann ist ihr Phasenunterschied am Aufpunkt in Richtung α durch die verschiedenen, bis zu $\Delta x = a \sin \alpha$ unterschiedlichen Laufwege bestimmt. Für $a \sin \alpha = \lambda$ ist der Wegunterschied von Punkten der oberen Spalthälfte und analogen Punkten der unteren Spalthälfte gerade $\lambda/2$. Das heißt, die Wellen löschen sich paarweise aus. In dieser Richtung wie auch für $a \sin \alpha = n\lambda$, $n = 1, 2, 3, \ldots$ tritt ein Minimum der Intensität auf.

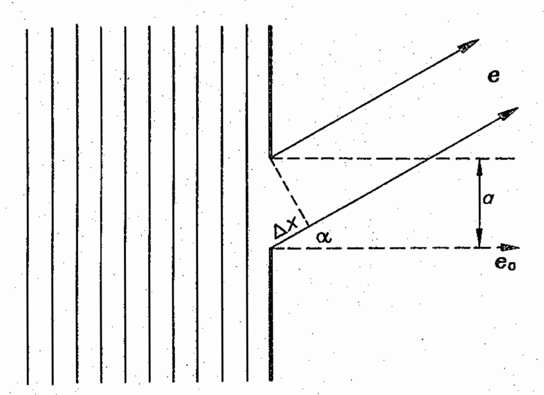


Abb. 2.3.4–4 Beugung an einem Spalt. Beobachtung der Intensität in Richtung e bei in Richtung e_0 einfallender ebener Welle

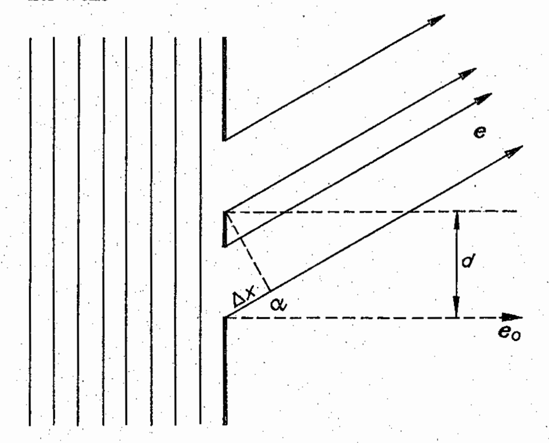


Abb. 2.3.4–5 Beugung am Doppelspalt

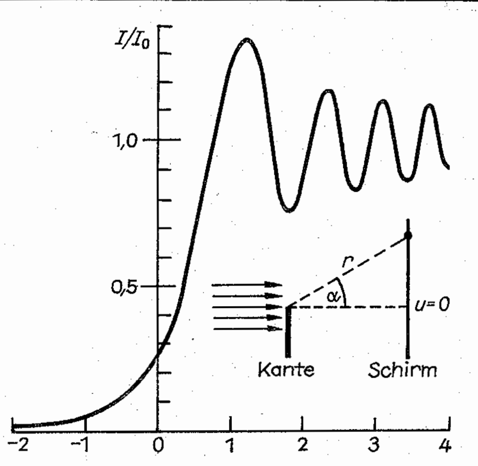


Abb. 2.3.4–6 Fresnelsche Beugung an einer Kante, $u = \sin 2\alpha \cdot \sqrt{r/(2\lambda)}$

Beugung am Doppelspalt (Abb. 2.3.4–5). Am Doppelspalt der Gitterkonstanten d haben die von analogen Punkten des oberen und unteren Spaltes ausgehenden Wellen einen Unterschied $\Delta x = d \sin \alpha$ im Laufweg. Sie haben einen Phasenunterschied π, wenn $d \sin \alpha = \lambda/2$ gilt. Minima und Maxima der Intensität treten bei den Winkeln $d \sin \alpha_{max} = n\lambda$, $d \sin \alpha_{min} = (n + 1/2)\lambda$ auf.

Aus der aus der Wellengleichung ableitbaren *Kirchhoffschen Formel*

$$u(\boldsymbol{r}) = \oiint_{\text{Rand}} \mathrm{d}a' \left\{ \frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} \frac{\partial}{\partial \boldsymbol{r}'} u(\boldsymbol{r}') - u(\boldsymbol{r}') \frac{\partial}{\partial \boldsymbol{r}'} \frac{\mathrm{e}^{\mathrm{i}k|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|} \right\}$$

kann man das Huygenssche Prinzip begründen. Diese Formel enthält die Aussage, daß man das Wellenfeld innerhalb eines geschlossenen Volumens berechnen kann, wenn man auf dem Rande die Wellenfunktion und deren Ableitung in Normalenrichtung (nach innen) kennt. Diese bestimmen die Amplitude und Phase vom Rande ausgehender Kugelwellen. Wenn man – z. B. im Falle des Spaltes (vgl. Abb. 2.3.4–4) – annimmt, daß die Wellenfunktion auf der Rückseite der Wand verschwindet und daß keine von rechts einfallenden Wellen auftreten, dann trägt zum Integral in der Kirchhoffschen Formel nur die Spaltöffnung bei. Bei der Auswertung ersetzt man dann noch die Amplitude der Wellenfront im Spalt durch den Wert, den sie ohne Wand haben würde.

Frauenhofersche Beugung. Damit bezeichnet man das Bild der Intensitätsverteilung in großem Abstand (groß gegen die Ausdehnung des beugenden Objektes) von den beugenden Objekten. Dort kann man die Kugelwellen durch ihren asymptotischen Wert ersetzen. Weiterhin soll paralleles Licht einfallen, d. h., der Abstand einer punktförmig gedachten Lichtquelle von dem beugenden Objekt soll auch groß gegen dessen Ausdehnung sein. Die Berechnungen für den Spalt und den Doppelspalt wurden in diesem Sinne durchgeführt.

Fresnelsche Beugung. Darunter versteht man das Beugungsbild im Nahfeld, in Abständen, die nicht groß gegen die Ausdehnung des beugenden Objektes bzw. gegen die Wellenlänge sind. Abb. 2.3.4–6 zeigt den Intensitätsverlauf

bei der *Beugung an der Kante* eines Schirmes. Auf der Lichtseite treten Intensitätsschwankungen auf. An der Schattengrenze ist die Intensität noch nicht auf Null abgeklungen.
Die Fresnelsche Beugung bestimmt das *Auflösungsvermögen* optischer Geräte, da das Licht an den Rändern der Blenden und an den Einfassungen der Linsen gebeugt wird. Wird paralleles Licht, z. B. bei der Betrachtung eines Fixsternes mit dem Fernrohr, von einer Linse (Radius R) im Brennpunkt (Brennweite f) gesammelt, dann entsteht kein heller Punkt, sondern eine *Beugungsscheibe* umgeben von dunklen und hellen Ringen. Der Kreis des ersten Minimums hat einen Radius $r = 0{,}61\,\lambda f/R$.

2.3.5. Schallwellen

Die allgemeinen Aussagen über die Wellenausbreitung treffen auch auf die Schallwellen zu. Vielfach wurde auch explizit auf Schallwellen Bezug genommen. In diesem Kapitel werden nur spezifische Probleme angesprochen.
2.3.5.1. Ausbreitung des Schalles. Schallwellen sind *mechanische Wellen*. Ihre Ausbreitung ist an ein Medium gebunden. Wenn man z.B. eine Klingel in einem Rezipienten an einen schlecht schalleitenden Faden aufhängt und die Luft auspumpt, sieht man die Bewegung des Klingelklöppels, ohne die Klingel zu hören. Läßt man Luft in den Rezipienten einströmen, dann ist die Klingel normal zu hören.
In *Gasen* und *Flüssigkeiten* breiten sich longitudinale Wellen aus, deren Ausbreitungsgeschwindigkeit $c^2 = \partial p/\partial \varrho$ aus der Zustandsgleichung folgt. Man muß beachten, daß die Dichteschwankungen adiabatisch verlaufen, d. h. kein Wärmeaustausch stattfindet. Damit enthält die *Schallgeschwindigkeit* in einem idealen Gas (pV^{γ} = konst.) den Adiabatenkoeffizienten $\gamma = c_p/c_v$, d. h. das Verhältnis der spezifischen Wärmen bei konstantem Druck bzw. konstantem Volumen. In Tab. 2.3.5–1 sind einige Ausbreitungsgeschwindigkeiten zusammengestellt.
In *Festkörpern* können longitudinale und transversale Wellen angeregt werden (vgl. 2.2.2. und 6.3.2.). Deren Ausbreitungsgeschwindigkeiten sind verschieden groß. Homogene, d. h. polykristalline Materialien, werden durch zwei elastische Konstanten charakterisiert, z. B. *Elastizitätsmodul* E und *Poissonsche Konstante* ν. Die Ausbreitungsgeschwindigkeiten hängen von ihnen in der Form

$$c_L = \sqrt{\frac{E(1-\nu)}{\varrho(1+\nu)(1-2\nu)}}, \quad c_T = \sqrt{\frac{E}{\varrho 2(1+\nu)}}$$

ab. Einige Werte sind in Tab. 2.3.5–2 enthalten.
In *Einkristallen* ist die Ausbreitungsgeschwindigkeit richtungsabhängig. Sie haben mehr als zwei voneinander unabhängige elastische Konstanten. Ihre Anzahl wird durch die Struktur bestimmt, bei hexagonalen Kristallen sind es fünf. Durch Messung verschiedener Ausbreitungsgeschwindigkeiten können sie gewonnen werden. Abb. 2.3.5–3 zeigt, wie durch verschiedene Anordnungen des schallgebenden Quarzes (schraffierte Kreise) longitudinale und transversale Ultraschallwellen – je nach der Polarisation der Schwingung des Quarzes – unterschiedlicher Ausbreitungsrichtung angeregt werden. Für Kalium-Barium-Nitrit $K_2Ba(NO_2)_4$ werden z. B. bei 150 °C die Ausbreitungsgeschwindigkeiten $v_{xL} = 3\,254$ m/s, $v_{xT} = 1\,676$ m/s, $v_{zL} = 4\,424$ m/s, $v_{zT} = 1\,687$ m/s und $v_{QL} = 3\,647$ m/s gemessen.

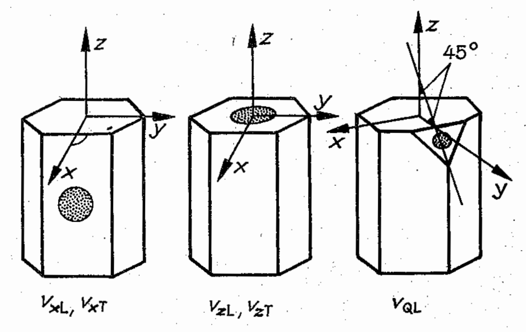


Abb. 2.3.5–3 Anregung fünf verschiedener Schallwellen in einem hexagonalen Einkristall

An *Grenzflächen* werden Schallwellen reflektiert (vgl. 2.3.4.1.). Bei bekannter Schallgeschwindigkeit können aus der Laufzeit eines Schallimpulses bis zum Eintreffen des Echos die Entfernung der Grenzfläche bestimmt (Echolot) bzw. Materialinhomogenitäten lokalisiert werden (*zerstörungsfreie Materialprüfung* mittels Ultraschall). Häufig wird dieses *Impulsechoverfahren* benutzt, um die Schallgeschwindigkeit zu ermitteln. Durch Verwendung von Ultraschallquellen können Messungen an Proben mit einer Ausdehnung im Zentimeterbereich durchgeführt werden.
Im Bereich des hörbaren Schalles treten starke *Beugungserscheinungen* auf. Bei einer Ausbreitungsgeschwindigkeit von 330 m/s gehört zu einem Ton der Frequenz 1 000 Hz eine Wellenlänge von 0,33 m. Die natürlichen Hindernisse haben also eine Ausdehnung von der Größenordnung der Wellenlänge des hörbaren Schalles. Daher kann man gut „um die Ecke hören".

Tab. 2.3.5–1 Schallgeschwindigkeiten bei 20 °C (nach LANDOLT-BÖRNSTEIN)

Substanz	c_s in m/s	Substanz	c_s in m/s
Aluminium	5 170	Eis (−4 °C)	3 230
Eisen	5 120	Wasser	1 490
Kupfer	3 700	Quecksilber	1 450
Beton	1 660	Luft	343
Kronglas	5 200	Kohlendioxid	268
Rotbuche	3 500	Stickstoff	348
Leinenschnur	1 800	Sauerstoff	324

Tab. 2.3.5–2 Elastische Konstanten und Ausbreitungsgeschwindigkeit longitudinaler und transversaler Wellen (nach TIETZ)

Substanz	E in 10^8 Pa/mm²	ν	ϱ in g/cm³	v_L in m/s	v_T in m/s
Aluminium	705	0,34	2,7	6 320	3 130
Silber	895	0,36	10,5	37,90	1 760
Gold	840	0,42	19,29	3 360	1 240
Kupfer	1 175	0,34	8,93	4 730	2 300
Eisen	2 080	0,29	7,86	5 900	3 230
Magnesium	493	0,28	1,74	5 900	3 280
Blei	235	0,41	11,3	21,60	700
Platin	1 670	0,30	21,4	3 260	1 730
Quarzglas			2,6	5 570	3 520

Die *Energiedichte* einer Schallwelle (vgl. 2.3.2.4.) ist eine Funktion von Ort und Zeit. Mittelt man sie für eine ebene Welle über den Bereich von einer Wellenlänge oder über die Zeit einer Schwingungsdauer, dann erhält man $w = (\varrho/2)v_{max}^2$. v_{max} ist der Maximalwert der *Schallschnelle*. Man kann ihn auch durch den maximalen Ausschlag y_{max} oder den Maximalwert p_{max} der Druckschwankung (*Schalldruck*) ausdrücken. Diese Größen hängen mit der Amplitude u_0 der Wellenfunktion über $y_{max} = u_0/c$, $v_{max} = \omega u_0/c = \omega y_{max}$ bzw. $p_{max} = \varrho\omega u_0 = \varrho c v_{max}$ zusammen. Manchmal verwendet man auch effektive Werte $v_{eff} = v_{max}/\sqrt{2}$, $p_{eff} = p_{max}/\sqrt{2}$. Dann gilt $w = \varrho v_{eff}^2$. Die Energiestromdichte $j = cw$ wird als *Schallintensität* bezeichnet. *Impulsstromdichte* und *Schallstrahlungsdruck* sind ihr proportional. In Tab. 2.3.5–4 sind diese Größen für ein Beispiel zusammengestellt.

Tab. 2.3.5–4 Schallgrößen für Luft und Wasser bei $f = 1\,000$ Hz und einer Schallintensität von 1 W/m² bei 20 °C

Größe	Formel	Luft	Wasser
Dichte	ϱ	1,189 kg/m³	998 kg/m³
Schallgeschwindigkeit	c	343 m/s	1 490 m/s
Frequenz	f	1 000 Hz	1 000 Hz
Intensität	j	1 W/m²	1 W/m²
Energiedichte	$w = j/c$	$2{,}92 \cdot 10^{-3}$ J/m³	$0{,}671 \cdot 10^{-3}$ J/m³
Schallschnelle	$v_{max} = \sqrt{2\,w/\varrho}$	0,070 ms⁻¹	$1{,}16 \cdot 10^{-3}$ ms⁻¹
Schalldruck	$p_{max} = \varrho c v_{max}$	28,5 Pa	1 720 Pa
Ausschlag	$y_{max} = v_{max}/2\pi f$	11 µm	0,18 µm

Da unsere Empfindungsstärke (vgl. 2.3.5.4.) logarithmisch mit der Schallintensität wächst, wählt man als Vergleichsmaß zwischen zwei Schallintensitäten den Logarithmus ihres Verhältnisses. Das *Dezibel* (dB) ist über $q = 10\lg(j_1/j_2)\,\text{dB} = 20\lg(p_1/p_2)\,\text{dB}$ definiert. Es wird dem dimensionslosen Vergleichswert q angefügt, um zu kennzeichnen, daß dieser über die angegebene Gleichung definiert wurde.

Tab. 2.3.5–5 Schallisolationsmaß (dB) (nach Landolt-Börnstein)

Material	128 Hz	512 Hz	2 048 Hz
Aluminium (0,63 mm)		13,2	23,2
Sperrholz (3,2 mm)	12	17	26
Gipsplatte (12,5 mm)	27	28	33
Haarfilz (25 mm)	4,9	6	6,7
Ziegelmauer (200 mm)		47,6	63,5
Holzbalkendecke	58	55	58
Glaswolle* (16 mm)		54	
Betonwand*		41	
Holztür*		34	
Fenster*		33	
Doppelfenster*		52	

*Mittelwerte über einen Frequenzbereich

Tab. 2.3.5–6 Schallabsorptionsgrad (nach Landolt-Börnstein)

Material	128 Hz	512 Hz	2 048 Hz
Sperrholz mit Tapete (11 mm)	0,10	0,06	0,07
Linoleum	0,02	0,03	0,04
Teppich auf Filz (10 mm)	0,11	0,37	0,27
Glas	0,035	0,027	0,02
Ziegelmauer (46 cm)	0,02	0,03	0,05
Haarfilz (25 cm)	0,12	0,51	0,60
Publikum im Theater	0,72	0,95	1,00
gepolsterte Theatersitze		0,42	

Schallwellen werden gedämpft. Die *Dämpfung* spielt in der Bau- und Raumakustik eine wesentliche Rolle. Das Dezibel wird auch als Maß für die *Schalldämmung* einer Wand benutzt. In q bedeutet dann j_1 die Intensität der einfallenden Schallwelle und j_2 die Intensität des durch die Wand hindurchgetretenen Anteiles. Tab. 2.3.5–5 enthält einige Beispiele für das Schallisolationsmaß q, und Tab. 2.3.5–6 führt einige Werte des Schallabsorptionsgrades A, $A = 1 - [(z - z_L)/(z + z_L)]^2$ (vgl. 2.3.4.1.) auf. Darin ist z_L der Wellenwiderstand der Luft. Soll der Schall von einer Wand „geschluckt" werden, dann muß A nahe Eins liegen.

2.3.5.2. Schallquellen. Die Erzeugung des Schallfeldes beruht darauf, daß bestimmte Körper oder Medien in Schwingungen versetzt werden. Die abgestrahlte Schalleistung ist in starkem Maße davon abhängig, wie gut die Kopplung zum umgebenden Medium ist. Schwingt ein anfangs angeregter Körper lange nach, z. B. eine angezupfte Saite, dann ist das ein Zeichen für eine schlechte Kopplung.

Eine eingespannte *Saite* erzeugt einen Grundton, dessen Wellenlänge durch die doppelte Saitenlänge $2L$ gegeben ist. Die Frequenz $f = (1/2L)\sqrt{\sigma/\varrho}$ steigt mit der Wurzel aus der Spannung σ an. Die Frequenzen der Obertöne sind ganzzahlige Vielfache der Grundfrequenz.

Eine *Stimmgabel* ist meist auf den *Kammerton* a' (440 Hz) geeicht. Da ihre beiden Arme entgegengesetzt schwingen, ist die abgestrahlte Intensität stark richtungsabhängig, wovon man sich durch Drehen der Stimmgabel überzeugen kann. Dies ist ein Interferenzeffekt zwischen den von den einzelnen Armen abgestrahlten Wellen.

Eine *Platte* oder eine *Membran* hat Eigenschwingungen, die durch Materialart, Form und Halterung bestimmt werden (vgl. 2.2.2.2.). Dabei sind die Frequenzen der Oberschwingungen keine ganzzahligen Vielfache der Grundfrequenz. Auf eine Platte gestreutes Korkmehl sammelt sich an den Schwingungsknoten an. Es entstehen die sogenannten *Chladnischen Klangfiguren*. In Pfeifen werden Luftsäulen in Schwingungen versetzt. Bei einer *offenen Pfeife*, d. h. bei einer beiderseitig offenen Luftsäule, befindet sich an den Pfeifenöffnungen ein Schwingungsbauch, die Wellenlänge des Grundtones ist etwa gleich der doppelten Pfeifenlänge. Bei einer *gedackten Pfeife*, d. h. bei einer nur einseitig offenen Luftsäule, ist die Wellenlänge des Grundtones etwa gleich dem Vierfachen der Pfeifenlänge. Der Wert der Wellenlänge wird durch Form und Durchmesser der Mündung beeinflußt. Bei den Obertönen einer Pfeife ist die Knotenzahl im Innern der Pfeife erhöht. Bei einer gedackten Pfeife wird bei starkem Hineinblasen der Grundton nur schwach angeregt. Bei der *Zungenpfeife* erfolgt die Erregung des Tones durch ein schwingendes elastisches Plättchen. Bei der

Lippenpfeife strömt die Luft gegen eine scharfe Schneide. In der turbulenten Strömung lösen sich periodisch Wirbel ab. Bei der *Lochsirene* wird ein Luftstrom auf eine rotierende mit Löchern versehene Scheibe geblasen.

Die *Ultraschallerzeugung* auf mechanischem Wege erfolgt mittels der Galtonpfeife oder der Ultraschall-Lochsirene. Die *Galtonpfeife* ist eine gedackte Pfeife mit veränderlicher Pfeifenlänge. Es können Schwingungen bis etwa 40 kHz erzeugt werden, die Schalleistung ist gering. Die *Ultraschall-Lochsirene* ist eine leistungsstarke Quelle, die abgestrahlte Leistung kann einige Kilowatt betragen. Sie wird z. B. zum Emulgieren eingesetzt.

Für die Erzeugung hochfrequenter Ultraschallwellen im MHz- und GHz-Bereich verwendet man vorwiegend *piezoelektrische Ultraschallgeber*. Piezoelektrische Kristalle, wie *Quarz, Seignettesalz, Zinkblende*, ändern unter dem Einfluß eines elektrischen Feldes ihre Dicke. Bringt man eine Probe zwischen zwei Kondensatorplatten und schließt ein Wechselfeld an, so werden mechanische Schwingungen mit der Frequenz des elektrischen Feldes erzeugt. Entspricht die Frequenz des elektrischen Feldes einer Eigenfrequenz des Quarzes, dann wird besonders viel Energie aus dem Feld aufgenommen, und man hat eine leistungsstarke Ultraschallquelle. Je nach der Orientierung des Kristalles lassen sich longitudinale oder transversale Wellen anregen.

2.3.5.4. **Hören.** Das menschliche Ohr kann Frequenzen im Bereich von 16 bis 20 000 Hz hören. Dieser Bereich ist individuell etwas verschieden, im Alter wird er schmaler. Das Ohr hat im empfindlichsten Bereich ein Auflösungsvermögen von $f/df \approx 300$, und es kann Lautstärken von einem Phon Unterschied auseinanderhalten. Daraus resultiert z. B. die Forderung, daß eine HiFi-Anlage im Bereich von 20 bis 16 000 Hz die Lautstärke proportional wiedergeben soll.

Während die *Schallintensität* oder *Reizstärke* eine physikalische Größe ist (vgl. 2.3.5.2.), charakterisiert die *Empfindungsstärke* den sinnlichen Reiz. Die gerade noch wahrnehmbare Empfindungsstärke bezeichnet man als *untere Hörschwelle*, die noch erträgliche Empfindungsstärke als *Schmerzgrenze* oder *obere Hörschwelle*. Die in Abb. 2.3.5–7 dargestellte *Hörfläche* zeigt, daß die zu diesen beiden Hörschwellen gehörige Schallintensität stark mit der Frequenz variiert; bei den tiefen Tönen gehört z. B. zur unteren Hörschwelle eine um einen Faktor 10^6 größere Schallintensität im Vergleich zum empfindlichsten Bereich.

Abb. 2.3.5–7 zeigt, daß im Empfindungsbereich des Ohres die Schallintensität um über zwölf Zehnerpotenzen variieren kann. Die Druckschwankungen in der Schallwelle überstreichen dabei sechs Zehnerpotenzen. Tab. 2.3.5–8 enthält einige Druckwerte. Die von uns empfundene *Lautstärke* wächst dabei keinesfalls proportional mit der Intensität an,

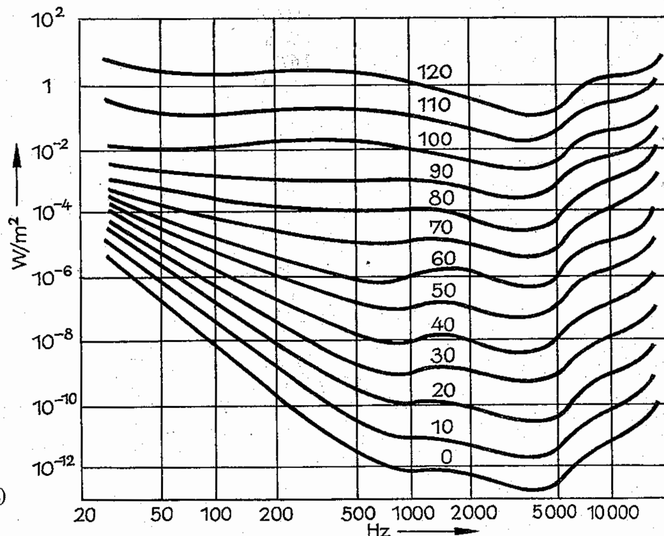


Abb. 2.3.5–7 Hörfläche (nach RECKNAGEL)

Auch *magnetostriktive Effekte* lassen sich zur Ultraschallerzeugung nutzen. Das Prinzip ist das gleiche wie bei der elektrostriktiven Anregung, nur daß die Längenänderung durch ein Magnetfeld verursacht wird. Die erzielten Leistungen sind wegen der geringen Längenänderungen jedoch klein, so daß die elektrostriktive Anregung bevorzugt wird.

Explosion bzw. *Detonation* werden nicht zur Schallerzeugung, sondern wegen der Druckwelle genutzt. Hier treten Schallwellen und Strömung gekoppelt auf, so daß die Ausbreitungsgeschwindigkeit über der Schallgeschwindigkeit liegt. Es können Druckwellen mit Drücken $p < 10^{10}$ N/m² erzeugt werden. Dies wird z. B. in der dynamischen Hochdruckphysik genutzt, um Werkstoffe auf eine Matrize zu pressen.

Tab. 2.3.5–8 Variationsbereich des Schalldruckes (nach LANDOLT-BÖRNSTEIN)

Schallquelle	Entfernung in m	p_{max} in Pa	p_{min} in Pa
Unterhaltung	1	0,0059	0,0059
Sopran	6	0,42	0,008
Baß	6	0,77	0,0013
Posaune	1	2,28	0,65
Trompete	1	5,4	0,86
Flöte	1	1,4	0,16
Klavier	3	0,26	0,02
Sirene	30,5	112	

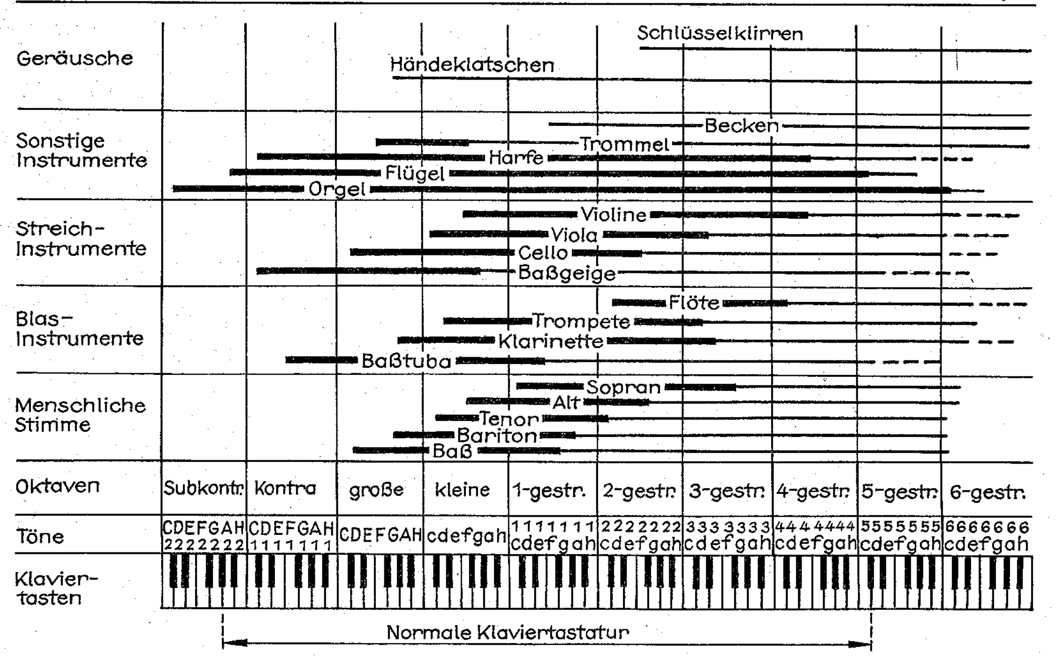


Abb. 2.3.5–9 Tonumfang verschiedener Instrumente. Der durch einen dünnen Strich markierte Bereich wird durch Obertöne erfaßt

nach dem *Weber-Fechnerschen Gesetz* wächst die Lautstärke etwa mit dem Logarithmus der Intensität an. Die Lautstärke wird in Phon angegeben. Für den Normalton von 1 000 Hz ist die Phonzahl durch $E = 10\lg(j/j_0)$ Phon gegeben. Darin bedeutet $j_0 = 10^{-12}\,\mathrm{W/m^2}$ etwa die Schallintensität an der unteren Hörgrenze bei 1 000 Hz. Für alle Frequenzen wird der Schallintensität, die als gleich laut empfunden wird, die gleiche Phonzahl zugeordnet. Die untere Hörschwelle hat also definitionsgemäß die Lautstärke 0 Phon. Die Schmerzgrenze liegt bei 120 bis 130 Phon. Entsprechend ihrer Definition ist die Phonzahl von zehn gleichen Schallquellen um 10 Phon höher als die Phonzahl der einzelnen Schallquelle. Die Lautstärke von z. B. einem Motorrad beträgt also noch etwa 90 % der Lautstärke von zehn Motorrädern. Als *Ton* wird die Empfindung bei einer reinen Schwingung bezeichnet. Als Bezugspunkt wird in der Musik vielfach der *Kammerton* a' = 440 Hz angenommen. Die *Oktave* ist das Intervall zwischen einem Ton und dem doppelter Frequenz. Das Klavier umfaßt z. B. die sieben Oktaven $_2A_1A$ A a a^1 a^2 a^3 a^4, d.h. einen Frequenzbereich von 27,5 bis 3 520 Hz (Abb. 2.3.5–9).

Tab. 2.3.5–10 Intervalle der Tonleiter (reine Stimmung)

Bezeichnung		f/f_0	Ganzzahlenreihe	f_{i+1}/f_i
Prime	c′	1	24	9/8
Sekunde	d′	9/8	27	
Terz	e′	5/4	30	10/9
Quarte	f′	4/3	32	16/15
Quinte	g′	3/2	36	9/8
Sexte	a′	5/3	40	10/9
Septime	h′	15/8	45	9/8
Oktave	c″	2	48	16/15

In der Musik unterteilt man eine Oktave üblicherweise in die in Tab. 2.3.5–10 aufgeführten sieben Intervalle. Bei einem Frequenzverhältnis 9/8 und 10/9 spricht man von einem *Ganztonschritt*, bei 16/15 von einem *Halbtonschritt*. Diese Aufteilung entspricht unserem Klangempfinden. Beginnt man die Tonleiter mit einem anderen Grundton als c^1, dann muß man Zwischentöne einschalten, um die gleichen Frequenzverhältnisse aufeinanderfolgender Töne zu erhalten. Beim Erhöhen entstehen die mit cis, dis, eis, fis, gis, ais und his bezeichneten Töne, beim Erniedrigen des, es, fes, ges, as, b und ces. Mit den vorgegebenen Frequenzverhältnissen, d.h. bei der reinen Stimmung, unterscheiden sich die Frequenzen der beiden Halbtonreihen etwas. Bei einem Instrument mit festen Tönen, wie z. B. beim Klavier, kann man daher die *reine Stimmung* nur für einen Grundton oder eine Tonart realisieren.

Da jedoch die Unterschiede gering sind, greift man zur *gleichteilig temperierten Stimmung*. Dabei bleibt im Vergleich mit der reinen Stimmung nur das Frequenzverhältnis der Oktave erhalten. Die Oktave wird in zwölf Halbtöne zerlegt, deren Frequenzverhältnis $f_{i+1}/f_i = \sqrt[12]{2} = 1{,}059\,5$ ist (Tab. 2.3.5–11). Dieser Wert bzw. sein Quadrat 1,122 5 ist mit 16/15 = 1,066 7 bzw. 9/8 = 1,125 oder 10/9 = 1,111 1 zu vergleichen, wenn man die Abweichungen von der reinen Stimmung sehen will.

Für die Musikempfindung ist neben *Lautstärke* und *Tonhöhe* die *Klangfarbe* eines Instrumentes maßgebend. Sie wird dadurch geprägt, daß *Obertöne* mit charakteristischer Intensität mitschwingen. Sie hängt davon ab, was zum Schwingen angeregt wird, Saiten (Streichinstrumente), Luftsäulen (Blasinstrumente), Membranen (Trommeln),

Tab. 2.3.5–11 Frequenzen bei gleichteilig temperierter Stimmung

Bezeichnung	f in Hz	Bezeichnung	f in Hz
c'	261,625	g'	391,996
cis'/des'	277,182	gis'/as'	415,305
d'	293,665	a'	440
dis'/es'	311,127	ais'/b'	466,164
e'	329,628	h'	493,883
f'	349,228	c''	523,250
fis'/ges'	369,994		

Platten (Becken) oder Stäbe (Xylophon). Bei einer Geige z. B. wird die Frequenz des Grundtones einer Saite durch deren Länge, Dicke und Spannung bestimmt. Die Obertöne haben ganzzahlige Vielfache dieser Frequenz. Sie bilden die *harmonische Tonreihe*. Wie Tab. 2.3.5–12 ausweist, enthält sie auch Töne, die in der Tonleiter nicht vorkommen. Für die Klangfarbe ist die relative Phasenlage von Grundton und Obertönen zueinander unwesentlich. Die zeitliche Ausbildung der Obertöne, d. h. der *Einschwingvorgang*, wie auch das Ausklingen sind jedoch für ein Instrument charakteristisch. So können mit elektronischen Musikinstrumenten verschiedene Musikinstrumente imitiert werden, da Stärke und zeitliches Zuschalten von Obertönen eingestellt werden können. Auch die *Anschlagform* bestimmt die Klangfarbe. So kann man eine Saite streichen, zupfen oder schlagen. Die Klangfarbe ist deutlich verschieden.

Die *Abstrahlung des Schalles* und damit die Dämpfung der Schwingung wird durch die Kopplung des schwingenden Teiles mit der Luft bestimmt. Beim Schwingen großer Flächen (Trommel, Gong) lassen sich gut große Lautstärken erzielen. Eine dünne Saite hat nur eine geringe Kopplung zur Luft. Man koppelt sie daher mit *Resonatoren*, die eine große Strahlungsfläche haben, wodurch der Ton lauter zu hören ist. Der Resonator erhöht aber nicht die Energie des schwingenden Systemes, er sorgt nur dafür, daß die vorhandene Energie rascher abgestrahlt wird. Der Resonator beeinflußt das Klangbild maßgeblich. Bei der Kopplung mit einer Saite werden bevorzugt die Schwingungen übertragen, für die Resonanz besteht, d.h. deren Frequenz mit einer Eigenfrequenz des Resonators übereinstimmt. So werden Schwingungen, deren Wellenlänge (bei Schwingungen des Resonatormateriales) größer als die Ausdehnung des Resonators ist, im Resonator weder merklich angeregt noch von ihm stark abgestrahlt.

In einem *Geräusch*, z.B. Meeresrauschen, Zerknittern eines Papiers, Händeklatschen, findet man ein kontinuierliches Spektrum von Frequenzen vor. Auch für die Frequenzabhängigkeit der Intensität gibt es keine definierten Verhältnisse.

2.3.6. Nichtlineare Wellen

Die in der Physik zu lösenden Gleichungen sind i. allg. nichtlinear. Vielfach können jedoch lineare Gleichungen diskutiert werden, weil für viele Anwendungen die zu berechnende Wellenfunktion klein ist, so daß quadratische Terme (oder höhere Ordnungen) vernachlässigt werden können.

Die Nichtlinearitäten entstehen in der Regel durch die *Materialgleichungen*. Untersucht man das Verhalten eines Plasmas, dann ist das Magnetfeld nicht allein von außen aufgeprägt, sondern es entsteht auch durch die Bewegung der geladenen Teilchen. In der Lorentz-Kraft $\boldsymbol{v} \times \boldsymbol{B}$ tritt dadurch ein in der Geschwindigkeit quadratischer Term auf. Die Wellengleichung für elektromagnetische Wellen in Substanzen ist nur dann linear, wenn Polarisation und Magnetisierung linear mit den Feldern zusammenhängen. Es ist jedoch bekannt, daß bei paramagnetischen und parelektrischen Substanzen diese lineare Abhängigkeit nur bei schwachen Feldern besteht. Auch die Navier-Stokessche Gleichung für die Strömung einer Flüssigkeit ist nichtlinear.

Die nichtlinearen Gleichungen sind wesentlich schwerer zu lösen als lineare. Aus Einzellösungen kann man nicht durch Überlagerung allgemeinere Lösungen aufbauen.

Nichtlineare Gleichungen haben auch stationäre Lösungen. Das sind solche, bei denen die Wellenfunktion nur von der Differenz $x - vt$ (im eindimensionalen Fall) abhängt. Diese Lösungen beschreiben eine Welle, die sich formunverändert ausbreitet. Es treten zwei Typen solcher Lösungen auf, periodische Wellen und lokalisierte Wellen, auch *Solitonen* genannt.

Die Nichtlinearität in der Wellengleichung wirkt auf eine Veränderung der Amplitude und Form der Wellenfunktion hin. Eine Formveränderung tritt jedoch auch durch Dispersion auf. Dies wurde schon bei den linearen Wellengleichungen gefunden. Außerdem bewirkt die Dämpfung eine Abnahme der Amplitude. Diese Effekte können sich so kompensieren, daß eine formunveränderte Ausbreitung stattfindet. Bei Stoßwellen hat man im wesentlichen ein Zusammenspiel von Nichtlinearität und Dämpfung. Wasserwellen sind ein Beispiel für periodische Wellen. Schon in linearer Näherung hängt ihre Ausbreitungsgeschwindigkeit wesentlich von den Randbedingungen ab. Im flachen Wasser (Tiefe h) ist die Dispersionsbeziehung $\omega(k) = k\sqrt{gh}$, d. h. die Ausbreitungsgeschwindigkeit $v = \sqrt{gh}$. Im tiefen Wasser erhält man $v = \sqrt{g/k}$. Bei den Lösungen der nichtlinearen Gleichung hängt die Ausbreitungsgeschwindigkeit von der Amplitude ab. Es gibt eine kritische Amplitude, bei der die Lösung instabil wird. Wasserwellen schlagen über, wie man es am Strand beobachtet, oder es tritt Turbulenz auf.

2.4. Elektrizitätslehre

Die für die Elektrizitätslehre grundlegenden Beobachtungen sind die Kräfte zwischen Ladungen, zwischen Stabmagneten und zwischen Strömen. Da elektrische Ströme durch bewegte Ladungen entstehen, deutet sich schon hier ein Zusammenhang zwischen elektrischen und magnetischen Erscheinungen an.

Der einfachste Ausgangspunkt für den Aufbau der Elektri-

Tab. 2.3.5–12 Harmonische Tonreihe von C ausgehend

f/f_0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	C	c	g	c'	e'	g'		c''	d''	e''		g''			h''	c'''

zitätslehre ist das Coulombsche Gesetz für die Kraft zwischen zwei Ladungen. Die Kräfte zwischen Ladungen spielen bei technischen Anwendungen jedoch eine untergeordnete Rolle, hier dominieren Eigenschaften von Strömen. Als die für die Elektrizitätslehre charakteristische Grundgröße hat man daher die Stromstärke eingeführt. Ihre Einheit ist das Ampere (A).
Die Grundgleichungen der Elektrizitätslehre oder Elektrodynamik sind die Maxwellschen Gleichungen, die den räumlichen und zeitlichen Verlauf des elektromagnetischen Feldes bestimmen. Sie enthalten Aussagen über die Quellen und Wirbel beider Felder, sind also partielle Differentialgleichungen, die besagen, daß die elektrischen Ladungen die Quellen des elektrischen Feldes sind und daß es keine magnetischen Ladungen gibt. Die differentielle bzw. integrale Form lautet

$$\operatorname{div} \boldsymbol{D} = \varrho \qquad \operatorname{div} \boldsymbol{B} = 0$$
$$\oiint \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{A} = Q \qquad \oiint \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{A} = 0.$$

Die Ströme und die zeitliche Änderung des elektrischen Feldes bilden die Wirbel des Magnetfeldes, die Wirbel des elektrischen Feldes werden allein durch die zeitliche Änderung des Magnetfeldes bestimmt.

$$\operatorname{rot} \boldsymbol{E} = -\partial \boldsymbol{B}/\partial t \qquad \operatorname{rot} \boldsymbol{H} = \boldsymbol{j} + \partial \boldsymbol{D}/\partial t$$
$$\oint \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{s} = \frac{\mathrm{d}}{\mathrm{d}t} \int \boldsymbol{B} \cdot \mathrm{d}\boldsymbol{A} \qquad \oint \boldsymbol{H} \mathrm{d}\boldsymbol{s} = I + \frac{\mathrm{d}}{\mathrm{d}t} \int \boldsymbol{D} \cdot \mathrm{d}\boldsymbol{A}.$$

Daneben benötigt man Materialgleichungen für die elektrische Polarisation $\boldsymbol{P}(\boldsymbol{E})$ bzw. die Magnetisierung $\boldsymbol{M}(\boldsymbol{H})$, die definitionsgemäß über $\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}$ bzw. $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$ den Zusammenhang zwischen den Feldern $\boldsymbol{E}$ und $\boldsymbol{D}$ bzw. $\boldsymbol{B}$ und $\boldsymbol{H}$ herstellen.
Wegen der komplizierten Form der Maxwellschen Gleichungen steht hier die Diskussion von Spezialfällen, wie Elektrostatik, Magnetostatik, Wechselstromkreise, im Vordergrund, in denen die experimentellen Erscheinungen und theoretischen Zusammenhänge relativ leicht überschaubar sind.

2.4.1. Elektrostatik

2.4.1.1. Die elektrische Ladung und ihr Feld. Elektrisch geladene Körper stoßen sich ab oder ziehen sich an, je nachdem, ob sie gleichnamig oder ungleichnamig geladen sind. Man unterscheidet deswegen zwei Arten elektrischer Ladung, die positive und die negative. Eine elektrische Ladung ist immer an eine Masse gebunden. Die Kraft zwischen zwei Ladungen Q_1 und Q_2 (vgl. 2.1.3.3.) wirkt in Richtung der Verbindungslinie und ist dem Produkt aus den beiden Ladungen direkt und dem Quadrat des Abstandes r zwischen ihnen umgekehrt proportional $F = Q_1Q_2/(4\pi\varepsilon_0 r^2)$. Sie wird als Coulomb-Kraft bezeichnet. Zwei Ladungen sind gleich, wenn eine dritte Ladung auf jede von ihnen im gleichen Abstand die gleiche Kraft ausübt. $\boldsymbol{F}/Q_2$ ist die von der Ladung Q_1 am Ort der Ladung Q_2 ausgeübte Feldwirkung oder ihre *elektrische Feldstärke* $\boldsymbol{E} = [Q_1/(4\pi\varepsilon_0 r^2)]\boldsymbol{r}/r$. Die von n Ladungen Q_1, Q_2, ... Q_n auf eine $(n+1)$te Ladung Q ausgeübte Kraft resultiert aus der vektoriellen Addition der einzelnen Kräfte $\boldsymbol{F} = \sum \boldsymbol{F}_i$. Sie ist der Ladung Q proportional. Sind statt der Punktladungen Q_1, Q_2 usw. flächenhaft oder räumlich verteilte Ladungen felderzeugend, sind die am Ort der Ladung Q zu addierenden Kräfte proportional zu $\sigma \mathrm{d}A/r^2$ bzw. zu $\varrho \mathrm{d}V/r^2$, wobei $\sigma = \mathrm{d}Q/\mathrm{d}A$ bzw. $\varrho = \mathrm{d}Q/\mathrm{d}V$ die felderzeugende *Ladungsdichte* auf einer Fläche A bzw. in einem Volumen V sind. Die auf die Ladung Q ausgeübte Kraft resultiert dann aus der Integration der Teilkräfte.
Mit den Basiseinheiten Sekunde (vgl. 2.1.1.2.) und Ampere (vgl. 2.4.3.1.) ist auch die Einheit der elektrischen Ladung Coulomb festgelegt; 1 C = 1 As.
Jede Ladung besteht aus kleinsten natürlichen Einheiten, die nicht weiter teilbar sind, den negativ geladenen *Elektronen* und den positiv geladenen *Protonen*. Ihre Ladung heißt *Elementarladung*. Diese Ladung ist dem Betrage nach $e = 1{,}602 \cdot 10^{-19}$ C.
Die elektrischen Ladungen unterliegen einem Erhaltungssatz: In einem abgeschlossenen System bleibt die Summe aus positiven und negativen Ladungen konstant. Die Aufladung eines Körpers geschieht in den meisten Fällen durch

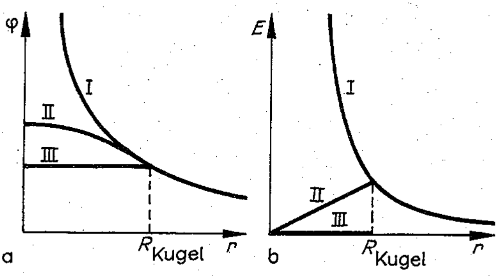
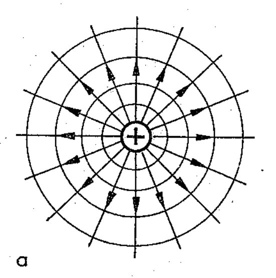
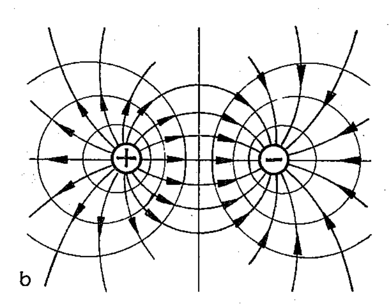


Abb. 2.4.1–2 (a) Potentialverlauf $\varphi(r)$ und (b) elektrische Feldstärke $\boldsymbol{E}(\boldsymbol{r})$ einer Punktladung (I) und einer homogenen, in einer Kugel verteilten Ladung (II) sowie einer gleichverteilten Oberflächenladung (III). Für $r > R$ haben φ und $\boldsymbol{E}$ für I, II, III denselben Verlauf





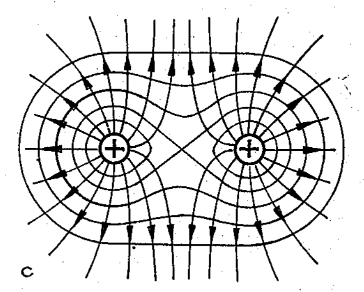


Abb. 2.4.1–1 Kraftlinien und Linien φ = konst. (a) einer punktfömigen positiven Ladung, (b) zwischen den Ladungen eines Dipols und (c) zwischen zwei gleichgroßen positiven Ladungen

Abgabe oder Aufnahme von Elektronen. Die Körper sind dann positiv bzw. negativ geladen. Geräte zur Messung von Ladungen heißen Elektrometer (vgl. 2.4.1.4.).
Ein elektrisches Feld wird auch mit Hilfe von *Kraftlinien* (vgl. 2.1.3.6.) beschrieben (Abb. 2.4.1–1). Im elektrostatischen Fall, in dem alle Ladungen ruhen, stellen die Kraftlinien Verbindungen von positiven und negativen Ladungen dar. Sie enden nie im freien Raum, in sich geschlossene Kraftlinien gibt es nicht; man sagt, das elektrostatische Feld ist *wirbelfrei*. Zwei Kraftlinien schneiden sich niemals. Die Richtung einer Kraftlinie im elektrischen Feld wird durch die Richtung der Kraft auf eine positive, möglichst kleine Probeladung bestimmt. Ist eine Ladung als punktförmig anzusehen und die entgegengesetzte Ladung sehr weit entfernt, so gehen von ihr geradlinige Kraftlinien aus mit konstanter Anzahl je Raumwinkel (vgl. Abb. 2.4.1–1). Außerhalb jeder kugelsymmetrischen Ladungsverteilung der Gesamtladung Q ist $\boldsymbol{E}$ gleich groß und gleich dem Feld einer Punktladung (Abb. 2.4.1–2).
Als *elektrischen Kraftfluß* durch eine Fläche A bezeichnet man die Größe $\int \boldsymbol{D}\mathrm{d}\boldsymbol{A} = \psi_e$, $\boldsymbol{D} = \varepsilon_0\boldsymbol{E}$. Legt man in ein elektrisches Feld eine in sich geschlossene Fläche, so ist der elektrische Kraftfluß gleich der von der Fläche A im Volumen V eingeschlossenen Ladung $\oiint_V \boldsymbol{D}\mathrm{d}\boldsymbol{A} = Q_V$. Diese Aussage und die Wirbelfreiheit bilden die Grundgleichungen des elektrostatischen Feldes.
Um in einem elektrischen Feld eine Ladung Q zu verschieben, ist gegen die elektrische Kraft $\boldsymbol{F} = Q\boldsymbol{E}$ eine Arbeit zu leisten, $W_{12} = -Q\int_1^2 \boldsymbol{E}\mathrm{d}\boldsymbol{s} = Q(\varphi_2 - \varphi_1)$. Diese Arbeit ist von der Form des Weges zwischen 1 und 2 unabhängig, zwischen den Punkten 1 und 2 besteht die Potentialdifferenz $\varphi_2 - \varphi_1 = U$. Setzt man das Potential im Unendlichen gleich Null, dann ist durch $\varphi = -\int_\infty^P \boldsymbol{E}\mathrm{d}\boldsymbol{s}$ das Potential φ im Punkt P festgelegt. $\varphi = f(x, y, z)$ ist eine skalare Ortsfunktion (vgl. 2.1.3.6.). Ihr negativer Gradient ist die elektrische Feldstärke, $\boldsymbol{E} = -\mathrm{grad}\,\varphi$. Das Potential φ wird in Joule/Coulomb, die Potentialdifferenz in Volt angegeben. Daraus folgt für die Feldstärke $\boldsymbol{E}$ die Einheit 1 N/C = 1 V/m.
Alle Punkte gleichen Potentials liegen auf einer *Äquipotentialfläche*. Die Äquipotentialflächen einer punktförmigen Ladung sind konzentrische Kugelflächen, in deren Mittelpunkt die Ladung ist. Die Kraftlinien des elektrischen Feldes stehen immer senkrecht auf den Äquipotentialflächen (vgl. Abb. 2.4.1–1). Das Potential einer Punktladung Q im Abstand r ist $\varphi = Q/4\pi\varepsilon_0 r$ (vgl. Abb. 2.4.1–2).
Um eine Ladung Q aus dem Unendlichen ($\varphi = 0$) an einen Ort $\boldsymbol{r}$ mit dem Potentialwert $\varphi(\boldsymbol{r})$ zu transportieren, ist die Arbeit $Q\varphi(\boldsymbol{r})$ erforderlich. Sie ist gleich der potentiellen Energie $W(\boldsymbol{r})$ der Ladung Q im elektrischen Potential φ. Zwei Ladungen Q_1, Q_2 im Abstand $|\boldsymbol{r}_1 - \boldsymbol{r}_2|$ haben die potentielle elektrische Energie $W = Q_1Q_2/4\pi\varepsilon_0|\boldsymbol{r}_1 - \boldsymbol{r}_2|$. Eine homogen geladene Kugel mit dem Radius R und der Gesamtladung Q hat die potentielle elektrische Energie $W = (3/5)Q^2/4\pi\varepsilon_0 R$.

2.4.1.2. Der elektrische Dipol im elektrischen Feld. Zwei ungleichnamige, aber dem Betrag nach gleich große Ladungen im Abstand l bilden einen elektrischen Dipol (Abb. 2.4.1–3) mit dem Dipolmoment $\boldsymbol{p} = Q\boldsymbol{l}$, wobei $\boldsymbol{l}$ vom negativen zum positiven Pol gerichtet ist. Sein elektrisches Potential ist für $r \gg l$, $\varphi = \boldsymbol{p}\cdot\boldsymbol{r}/4\pi\varepsilon_0 r^3 = p\cos\alpha/4\pi\varepsilon_0 r^2$, sein elektrisches Feld $\boldsymbol{E} = [3(\boldsymbol{p}\cdot\boldsymbol{e})\boldsymbol{e} - \boldsymbol{p}]/4\pi\varepsilon_0 r^3$, $\boldsymbol{e} = \boldsymbol{r}/r$ (vgl. Abb. 2.4.1–1b). Im elektrischen Feld $\boldsymbol{E}$ greift an den beiden Ladungen eines Dipols ein Kräftepaar (vgl. 2.1.1.4.) an, so daß er ein Drehmoment $\boldsymbol{M} = \boldsymbol{p}\times\boldsymbol{E}$ erfährt (Abb. 2.4.1–4). In einem inhomogenen elektrischen Feld erfährt ein Dipol außer dem Drehmoment zusätzlich eine Kraftwirkung, $\boldsymbol{F} = \mathrm{grad}\,(\boldsymbol{p}\cdot\boldsymbol{E})$.

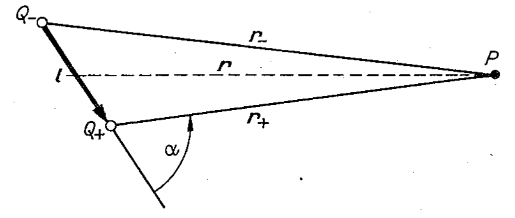


Abb. 2.4.1–3 Zur Berechnung des Potentials eines Dipols $r_- - r_+ \approx \boldsymbol{l}\cdot\boldsymbol{r}/r$

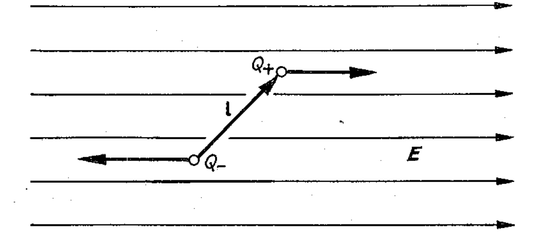


Abb. 2.4.1–4 Im homogenen elektrischen Feld an einem Dipol angreifendes Kräftepaar

2.4.1.3. Elektrische Ladung auf Leitern. Im Leiter sind elektrische Ladungen frei beweglich. Im elektrostatischen Fall befinden sich daher bei Aufladung alle Ladungen auf der Oberfläche, und $\boldsymbol{E}$ verschwindet im Innern. Die Kraftlinien des elektrischen Feldes stehen auf der Leiteroberfläche senkrecht. Das Potential hat im Innern und auf der Oberfläche eines Leiters den gleichen Wert, φ = konst.
Von zwei leitend miteinander verbundenen Metallkugeln trägt bei einer Aufladung die mit dem kleineren Radius wegen des gleichen Potentials auch die kleinere Ladungsmenge. Die Ladungsdichte σ auf ihrer Oberfläche ist aber im Verhältnis der beiden Radien größer: $Q_1/R_1 = Q_2/R_2$; $\sigma_1 = Q_1/4\pi R_1^2$; $\sigma_1/\sigma_2 = R_2/R_1$. D.h., je kleiner der Krümmungsradius der Oberfläche eines metallischen Körpers ist, um so größer ist bei einer Aufladung die Ladungsdichte auf seiner Oberfläche und damit die Anzahl der auf diesem Teil der Oberfläche endenden Kraftlinien des elektrischen Feldes im Außenraum. An Metallspitzen kann wegen der großen Krümmung (kleiner Krümmungsradius) bei einer Aufladung die Feldstärke so groß werden, daß es dort zu einem elektrischen Durchbruch, zu einer Gasentladung in Form einer sogenannten Büschel- oder Spitzenentladung kommt.
Im Innern eines metallischen Hohlkörpers, der keine Ladungen umschließt, ist die elektrische Feldstärke Null wie im massiven Metallkörper. Man kann einen Metallkörper vollständig entladen, wenn man ihn in das Innere eines metallischen Hohlkörpers ein- und mit dem Hohlkörper in Kontakt bringt. Bei einem so verwendeten metallischen Hohlkörper spricht man auch von einem Faraday-Becher

(Abb. 2.4.1–5). Ein Raum kann gegen statische elektrische Felder dadurch abgeschirmt werden, daß man ihn mit metallischen Wänden umgibt. In vielen Fällen erzielt man die gleiche Wirkung mit einem hinreichend engmaschigen metallischen Netz (*Faraday-Käfig*).

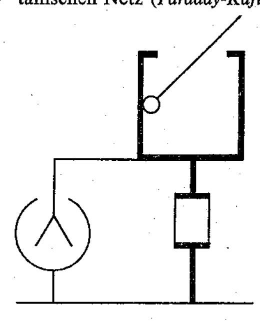


Abb. 2.4.1–5 Faraday-Becher mit Elektrometer zum Nachweis der aufgenommenen Ladung

Kapazität. Ein durch Aufladung auf ein bestimmtes Potential gebrachter isoliert aufgestellter Leiter trägt eine seinem Potential φ proportionale Ladung Q. Der Quotient Q/φ wird als seine Kapazität C bezeichnet. Die Einheit der Kapazität ist das Farad, $1\,\mathrm{F} = 1\,\mathrm{C}/1\,\mathrm{V} = 1\,\mathrm{As}/1\,\mathrm{V}$. Jedes Elektrometer hat eine bestimmte Kapazität, d. h., die von ihm angezeigte Ladung Q zeigt gleichzeitig eine Potentialdifferenz oder Spannung U an. Eine einzelne Kugel mit dem Radius R hat die Kapazität $4\pi\varepsilon R$.

Energie eines geladenen Leiters. Zur Heranführung einer Ladung $\mathrm{d}Q$ an einen Leiter mit der Ladung Q ist die Arbeit $\mathrm{d}W = U\mathrm{d}Q = Q\mathrm{d}Q/C$ erforderlich, wenn dabei die Potentialdifferenz U zu durchlaufen ist. Die gesamte Arbeit zur

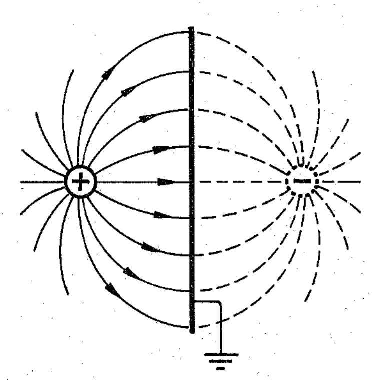


Abb. 2.4.1–6 Kraftlinien zwischen einer positiven Ladung und einer ebenen geerdeten Metallplatte

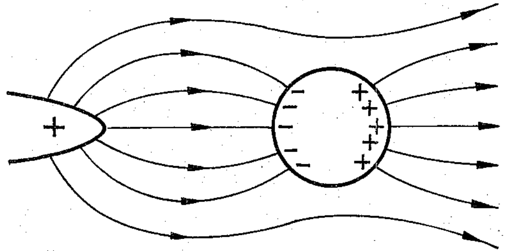


Abb. 2.4.1–7 Ladungstrennung auf einem ungeladenen metallischen Körper im elektrischen Feld durch Influenz

Aufladung eines Leiters ergibt sich daher zu $W = \int_0^Q Q\mathrm{d}Q/C = Q^2/2C = CU^2/2$.

Influenz. Bringt man in Abb. 2.4.1–1b eine ebene geerdete Metallplatte an die Stelle der Potentialfläche in der Mitte zwischen die beiden Ladungen, so ist das Kraftlinienbild zwischen positiver Ladung und Metallplatte unverändert (Abb. 2.4.1–6). Die Metallplatte trägt eine durch die positive Ladung influenzierte negative Ladung, deren Ladungsdichte der Kraftliniendichte am Ort der Metallplatte entspricht. Die Kraftlinien des elektrischen Feldes beginnen bei der positiven Ladung und enden auf der Metallplatte, also an dort befindlichen negativen Ladungen. Die auf der Platte feststellbare Flächenladungsdichte (Ladung/Fläche) $\sigma = \varepsilon_0 E$ ist in gleicher Weise ein Maß für das elektrische Feld. Es wird die elektrische Verschiebung $\boldsymbol{D} = \varepsilon_0 \boldsymbol{E}$, auch elektrische Flußdichte $\boldsymbol{D}$ genannt, eingeführt. Der Proportionalitätsfaktor ist die Dielektrizitätskonstante oder *Permittivität* des Vakuums, gekennzeichnet durch den Index 0. Sein Zahlenwert ist $\varepsilon_0 = 8{,}854\,187\,82 \cdot 10^{-12}\,\mathrm{As/Vm}$.

Entsprechend wird durch Influenz die Oberfläche eines isolierten und ungeladenen Metallkörpers in einem elektrischen Feld zu einem Ladungsträger, obwohl die Gesamtladung Null bleibt. Die Feldlinien enden an den negativen und beginnen an den positiven Ladungen (Abb. 2.4.1–7). Besteht der Metallkörper aus zwei in Kontakt befindlichen Metallplatten, dann trennt man mit den Platten auch die auf ihrer Oberfläche durch Influenz getrennten Ladungen.

Ein *Kondensator* besteht aus zwei gegeneinander isolierten Metallflächen. Bei einem Kugelkondensator (Abb. 2.4.1–8) sind die beiden Flächen konzentrische Kugelflächen. Das bei Aufladung sich zwischen den konzentrischen Kugelflächen ausbildende elektrische Feld ist kugelsymmetrisch und gleich dem einer Punktladung (vgl. Abb. 2.4.1–2). Die Spannung zwischen den Kugelflächen ist

$$U = -\int_{R_2}^{R_1} Q/(4\pi\varepsilon_0 r^2)\,\mathrm{d}r = Q/[4\pi\varepsilon_0 R_1 R_2/(R_2 - R_1)].$$

Für $R_1 \gg R_2 - R_1 = d$ darf man $R_1 \approx R_2 = R$ setzen. Damit folgt für die Kapazität $C_K = 4\pi\varepsilon_0 R^2/d = \varepsilon_0 A/d$. Sie ist der Größe der einander gegenüberliegenden Flächen A direkt und deren Abstand d umgekehrt proportional. Bei einem Plattenkondensator ist das elektrische Feld zwischen den Platten homogen bis auf ein gewisses Randgebiet (Abb. 2.4.1–9). Die zwischen den Platten herrschende Feldstärke ist der angelegten Spannung proportional $|\boldsymbol{E}| = U/d$. Bei Parallelschaltung zweier Kondensatoren addieren sich deren Kapazitäten, $C = C_1 + C_2$. Für Hintereinanderschaltung gilt dann $1/C = 1/C_1 + 1/C_2$. Da die einander gegenüberliegenden Platten eines Kondensators entgegengesetzte Ladung tragen, üben sie aufeinander eine anziehende Kraftwirkung aus, $|\boldsymbol{F}|/A = (1/2)\,\sigma|\boldsymbol{E}| = (1/2)\,\varepsilon_0 U^2/d^2$.

Elektrometer. Beim Einfadenelektrometer (Abb. 2.4.1–10) wird die zu messende Ladung auf einem isoliert befestigten und elastisch gespannten Metallfaden aufgebracht. Dieser befindet sich im elektrischen Feld von zwei auf entgegengesetztes Potential gebrachten Metallschneiden. Die Auslenkung des Fadens durch die an der Ladung angreifende Kraft $\boldsymbol{F} \sim Q\boldsymbol{E}$ wird mit einem Mikroskop beobachtet oder gemeinsam mit einer Skale auf einem Projektionsschirm abgebildet. Mit solchen Elektrometern können Ladungen von einigen $10^{-14}\,\mathrm{C}$ an aufwärts und umgerechnet über die Kapazität der Elektrometer Spannungen von einigen $10^{-3}\,\mathrm{V}$ an aufwärts gemessen werden.

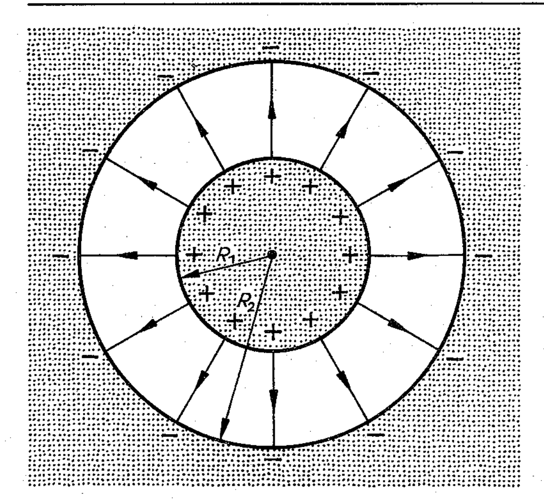


Abb. 2.4.1–8 Kugelkondensator

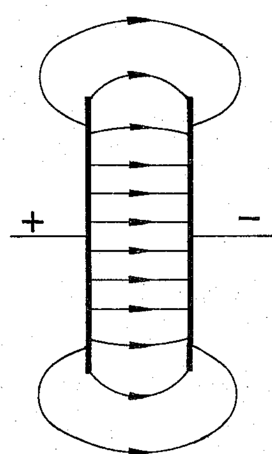


Abb. 2.4.1–9 Das elektrische Feld zwischen den Platten eines Kondensators mit seinem inhomogenen Randgebiet

2.4.1.4. Energiedichte des elektrischen Feldes. Beim Transport einer kleinen positiven Ladung dQ von der negativen zur positiven Platte eines Kondensators, dessen elektrisches Feld $E = U/d$ als homogen vorausgesetzt wird, ist die Arbeit $U\mathrm{d}Q$ erforderlich. Nach dem Transport dQ hat auch die Flächenladung und damit die elektrische Flußdichte zugenommen, $\mathrm{d}D = \mathrm{d}Q/A$. Die verrichtete Arbeit $U\mathrm{d}Q = E\mathrm{d}A\mathrm{d}D = VE\mathrm{d}D$ äußert sich in einer Zunahme der Feldenergie. Sie beträgt je Volumeneinheit $E\mathrm{d}\boldsymbol{D}$. Durch Integration erhält man die gesamte im elektrischen Feld gespeicherte Energie je Volumeneinheit zu

$$\int_0^Q U\mathrm{d}Q = \int_0^E E\varepsilon_0\,\mathrm{d}E = (1/2)\varepsilon_0 E^2 = (1/2)\boldsymbol{ED}.$$

2.4.1.5. Elektrische Eigenschaften der Stoffe. Bringt man zwischen die Platten eines aufgeladenen Kondensators einen Isolator (Dielektrikum) ein, so sinkt die Spannung. Bei gleicher Spannung ist die Kondensatorladung nach Einbringen eines Dielektrikums größer als ohne ein solches. Aus $Q = CU$ folgt, daß durch das Einbringen eines Dielektrikums die Kapazität des Kondensators erhöht wird. Als relative Dielektrizitätskonstante ε_r bezeichnet man den Zahlenfaktor, um den die Kapazität eines Kondensators durch Einbringen eines Dielektrikums erhöht wird. Man ordnet ihn dem Dielektrikum als eine Materialkonstante zu. Für die Kapazität eines Plattenkondensators gilt dann allgemein $C = \varepsilon A/d$ mit $\varepsilon = \varepsilon_r \cdot \varepsilon_0$. Das Absinken der Spannung hat seine Ursache in einer *Polarisation* $\boldsymbol{P}$ des Dielektrikums (vgl. 6.7.2.3.), der Erzeugung einer Dipoldichte (Abb. 2.4.1–11).

Von der Ladungsdichte ist auf den Platten der Anteil abzuziehen, der durch Polarisation gebunden wird. Dadurch vermindert sich die elektrische Feldstärke $\boldsymbol{D} - \boldsymbol{P} = \varepsilon_0 \boldsymbol{E}$. Durch Umformen erhält man $\boldsymbol{P} = \varepsilon_0 \chi_e \boldsymbol{E}$ mit der elektrischen *Suszeptibilität* des Dielektrikums $\chi_e = \varepsilon_r - 1$. $\boldsymbol{P}$ ist das aus den atomaren bzw. molekularen Dipolmomenten aufsummierte Dipolmoment je Volumeneinheit.

Kann man – wie in Gasen oder in verdünnten Lösungen – die Wechselwirkung zwischen den Atomen bzw. Molekülen vernachlässigen, dann ist die elektrische Feldstärke E_M, die ein Molekül spürt, gleich der äußeren Feldstärke E. Dann gilt $\varepsilon_0 \chi_e = N\alpha$ (α = molekulare Polarisierbarkeit, N = Anzahl der Moleküle je Volumen). Bei Berücksichtigung der

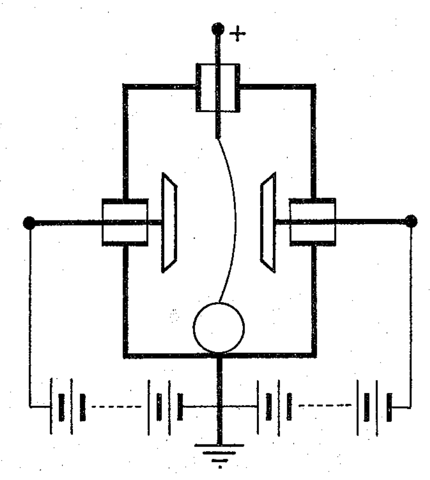


Abb. 2.4.1–10 Einfadenelektrometer

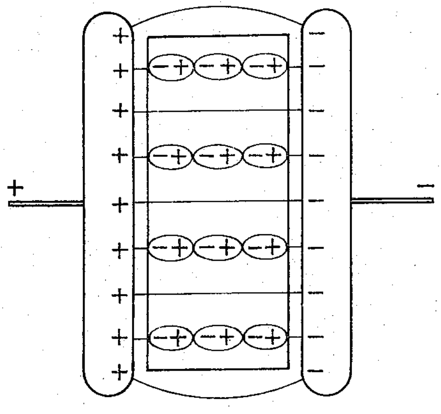


Abb. 2.4.1–11 Polarisation des Dielektrikums im elektrischen Feld eines Kondensators

Wechselwirkung zwischen den Molekülen, was in unpolaren Flüssigkeiten und Festkörpern erforderlich ist, gilt für die Feldstärke am Ort des Moleküls $\boldsymbol{E}_M = \boldsymbol{E} + \boldsymbol{P}/3\varepsilon_0$. Damit ergibt sich als Zusammenhang zwischen der elektrischen Suszeptibilität χ_e und der molekularen Polarisierbarkeit eines Stoffes die sogenannte *Clausius-Mosottische Gleichung* $\chi_e = (N\alpha/\varepsilon_0)/(1 - N\alpha/3\varepsilon_0)$. Über $\alpha = (3\varepsilon_0/L)(M/\varrho)(\varepsilon_r - 1)/(\varepsilon_r + 2)$ kann man α durch Messung von ε_r bestimmen, wobei $N = L\varrho/M$ verwendet wurde (vgl. 2.6.1.6.). Für das unpolare Wasserstoffgas H_2 beträgt z. B. die molekulare Polarisierbarkeit $\alpha = 1{,}95 \cdot 10^{-40}\,\mathrm{Asm^2/V}$, für die unpolare Flüssigkeit CCl_4 ist $\alpha = 11{,}7 \cdot 10^{-40}\,\mathrm{Asm^2/V}$.

Spontane elektrische Polarisation. So wie bei einem magnetischen Stoff nach Abschalten eines äußeren magnetischen Feldes ein Restmagnetismus verbleibt (vgl. 2.4.2.5.), gibt es den analogen Fall bei der elektrischen Polarisation. In Analogie zu den Ferromagnetika werden solche Stoffe *Ferroelektrika* genannt. In makroskopischen Kristallbereichen (Domänen) sind die durch die Kristallstruktur bedingten elektrischen Dipole einheitlich orientiert. Ferroelektrika gehen nach Überschreiten einer bestimmten stoffspezifischen Temperatur T_C (Curie-Punkt) in den *paraelektrischen Zustand* über. In ihm gehorcht die Suszeptibilität des Stoffes dem *Curie-Weißschen Gesetz*, $\chi_e = C'/(T - T_C)$. C' ist die sogenannte Curie-Konstante.

Kraftwirkung auf Dielektrika im elektrischen Feld. In einem homogenen elektrischen Feld orientiert sich ein frei beweglicher dielektrischer Körper so, daß das durch Polarisation in ihm erzeugte Dipolmoment am größten wird und in Richtung des Feldes zeigt. Ist er länger als dick und ist die Dielektrizitätskonstante ε des ihn umgebenden Mediums kleiner als seine eigene, stellt er sich mit seiner Längsachse in Richtung des Feldes, andernfalls quer dazu. In einem inhomogenen elektrischen Feld wird ein dielektrischer Körper in Bereiche größerer Feldstärke gezogen, wenn seine Dielektrizitätskonstante größer als die des ihn umgebenden Mediums ist, andernfalls aber in Bereiche geringerer Feldstärke abgedrängt. Zwischen den Platten eines senkrecht in ein flüssiges Dielektrikum eingetauchten Kondensators steigt bei Anlegen einer Spannung die Flüssigkeit gegen die Schwerkraft hoch (Abb. 2.4.1–12). Die Steighöhe ist $h = (1/\varrho g)(1/2)\,ED\,(\varepsilon_r - 1)$. Da ϱgh gleich dem hydrostatischen Druck der Flüssigkeitssäule ist, läßt sich das Ergebnis auch so deuten, daß an der Oberfläche eines flüssigen Dielektrikums ein tangentiales elektrisches Feld einen nach außen gerichteten Druck vom Betrag $(1/2)ED\,(\varepsilon_r - 1)$ bewirkt.

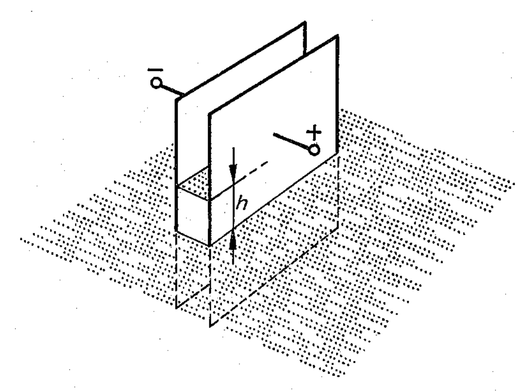


Abb. 2.4.1–12 Flüssiges Dielektrikum im elektrischen Feld eines Kondensators

Mit der Polarisation eines Dielektrikums kommt es auch zu einer *Elektrostriktion* in Form einer elastischen Spannung mit Form- und Volumenänderung.

Der *piezoelektrische Effekt* bewirkt das Auftreten von Ladungen auf den Grenzflächen eines Kristalles bei mechanischer Deformation (vgl. 6.7.3.).

Polarisation und elektrische Doppelschichten in der Oberfläche von festen und flüssigen Stoffen. Werden zwei verschiedene ungeladene feste Isolierstoffe miteinander gerieben, tragen die beiden Oberflächen nach dem Trennen entgegengesetzte Ladung. Die *Coehnsche Regel* besagt, daß der Stoff mit der größeren Permittivität nach dem Trennen positiv geladen ist. Es läßt sich eine Spannungsreihe der Stoffe derart aufstellen, daß bei Kontakt mit dem in der Reihe folgenden Stoff eine positive, bei Kontakt mit dem in der Reihe vorangehenden Stoff eine negative Aufladung erfolgt. Die Coehnsche Regel gilt auch für den Kontakt eines festen Stoffes mit einer nichtleitenden Flüssigkeit. Die *Wasserfall-* oder *Balloelektrizität* hat ihre Ursache im Vorhandensein einer elektrischen Doppelschicht an der Oberfläche der Wassertropfen, deren negativer Anteil nach außen gekehrt ist. Beim Fallen der Wassertropfen werden von der vorbeiströmenden Luft Partikeln in der Größenordnung von $10^{-6}\,\mathrm{cm}$ Durchmesser abgerissen, die dann als negative Wasserfallelektrizität festgestellt werden. Die eigentlichen, relativ großen Wassertropfen sind positiv geladen. Bei wässrigen Elektrolytlösungen ist der Effekt schwächer. An der Entstehung von Gewitterwolken ist die Balloelektrizität in entsprechender Weise beteiligt.

2.4.2. Magnetostatik

2.4.2.1. Der magnetische Dipol im magnetischen Feld. Das Verhalten einer Magnetnadel im Magnetfeld der Erde gleicht dem eines elektrischen Dipols in einem homogenen elektrischen Feld. Sie stellt sich – frei beweglich über ihren Schwerpunkt aufgehängt – in eine Nord-Süd-Richtung ein. Ihr nach Norden zeigendes Ende wird Nordpol, das andere Südpol genannt. Die Abweichung gegenüber geografisch Nord ist ihre Deklination. Die Abweichung gegen die Horizontale ist ihre Inklination. Die Magnetnadel zeigt die Richtung des Magnetfeldes der Erde an. Es weist vom Süd- zum Nordpol der anzeigenden Nadel.

Am Verhalten zweier Magnetnadeln stellt man fest: Gleichnamige Pole stoßen sich ab, ungleichnamige ziehen sich an. Im Gegensatz zum elektrischen Fall lassen sich in einem magnetischen Dipol polar entgegengesetzte magnetische Mengen nicht isolieren. Es läßt sich nur das magnetische Moment $\boldsymbol{m}$ eines Stabmagneten ermitteln. In einem homogenen Magnetfeld $\boldsymbol{H}$, $\boldsymbol{B} = \mu_0\boldsymbol{H}$, erfährt er ein *Drehmoment* $\boldsymbol{m} \times \boldsymbol{B}$. In einem inhomogenen Magnetfeld wirkt auf ihn außerdem die *Kraft* $\boldsymbol{F} = \operatorname{grad}(\boldsymbol{m} \cdot \boldsymbol{B})$, wobei $W = -\boldsymbol{m} \cdot \boldsymbol{B}$ die *Wechselwirkungsenergie* des Dipols im Magnetfeld ist. Ein Dipol führt, frei beweglich im Magnetfeld aufgehängt, bei kleinen Auslenkungen harmonische Drehschwingungen aus mit einer Schwingungsdauer $T = 2\pi\sqrt{J/mB}$. J ist das Trägheitsmoment des Dipols. Daraus folgt bei bekanntem T, J, H das magnetische Moment m, seine Einheit ist $\mathrm{Am^2}$.

Ein magnetischer Dipol erzeugt ein Magnetfeld $\boldsymbol{H}_D = [3(\boldsymbol{m} \cdot \boldsymbol{e})\,\boldsymbol{e} - \boldsymbol{m}]/(4\pi r^3)$, dessen Feldlinien analog zum elektrischen Fall (vgl. Abb. 2.4.1–1b) verlaufen. Um die

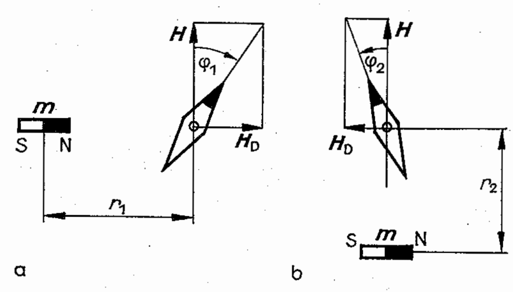


Abb. 2.4.2–1 Gaußsche Methode der Magnetfeldmessung (a) erste Hauptlage, (b) zweite Hauptlage

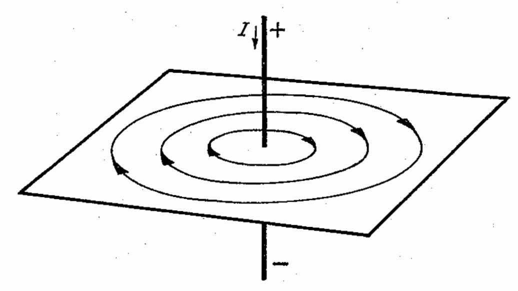


Abb. 2.4.2–4 Feldlinien des Magnetfeldes $\boldsymbol{H}$ um einen Stromleiter

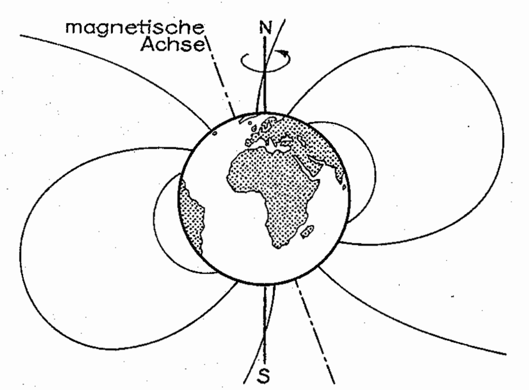


Abb. 2.4.2–2 Erdkugel mit magnetischen Kraftlinien (nach GRIMSEHL)

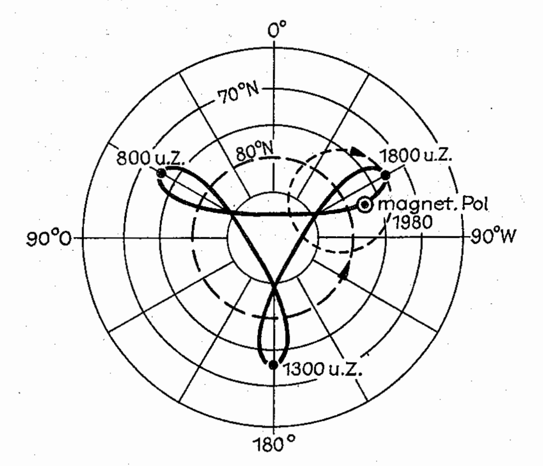


Abb. 2.4.2–3 Wanderung des magnetischen Pols der Erde (nach KAWAI et al.)

Stärke eines homogenen Magnetfeldes $\boldsymbol{H}$ zu messen, bestimmt man aus dem oben erläuterten Schwingungsversuch das Produkt mH. Weiterhin beobachtet man die Auslenkung einer Magnetnadel φ, wenn dem Feld $\boldsymbol{H}$ das Feld des Stabmagneten $\boldsymbol{H}_D$ überlagert wird. In der ersten bzw. zweiten *Gaußschen Hauptlage* (Abb. 2.4.2–1) stehen beide Felder senkrecht zueinander, daher gilt $\tan\varphi = H/H_D$ mit $H_{D1} = 2m/4\pi r_1^3$ *und* $H_{D2} = m/(4\pi r_2^3)$. Hier kann also das Verhältnis H/m bestimmt werden. Da T, J, r und jeder der beiden Winkel φ sehr genau meßbar sind, wird diese von GAUSS eingeführte Methode zur Bestimmung des magnetischen Momentes $\boldsymbol{m}$ von magnetischen Etalons und zur Absolutbestimmung der Horizontalkomponente des Erdmagnetfeldes auch heute noch angewendet.

2.4.2.2. Das Magnetfeld der Erde. Es ist in Erdnähe in erster Näherung gleich dem Feld einer magnetisierten Kugel (Abb. 2.4.2–2). Seine magnetische Achse fällt nahezu mit der Drehachse der Erde zusammen. Es hat seinen Ursprung in Konvektionsströmen des äußeren flüssigen Erdkernes (3000 bis 5000 km Tiefe), die durch *Coriolis-* und *Lorentz-Kraft* zu Wirbeln geschlossen werden. Trotz der zu erwartenden Rotationssymmetrie dieser Bewegungen wandert der magnetische Nordpol der Erde, der gegenwärtig bei der Bathurst-Insel im nördlichen Kanada liegt, im Laufe der Jahrhunderte merklich. Seine Bewegung kann als Überlagerung zweier Kreisbewegungen ähnlich einer Hypozykloide dargestellt werden, die im Laufe von etwa 1500 Jahren einmal durchlaufen wird (Abb. 2.4.2–3). Dieser säkularen Drift des magnetischen Poles der Erde sind relativ kurzzeitige Schwankungen, die aber bisher selten 20° überschritten haben, überlagert. Messungen des Restmagnetismus in ozeanischen Krustengesteinen – bis zu 80 Millionen Jahren zurückverfolgbar – lassen darauf schließen, daß im Mittel alle 200000 bis 300000 Jahre unregelmäßig und bisher noch mit keiner Theorie vorhersagbar Umpolungen des Erdmagnetfeldes stattgefunden haben. Das Magnetfeld der Erde, wie es als Dipolfeld verfolgbar ist, reicht bis zu 10 Erdradien in den Weltraum. Es ist die Ursache für die Strahlungsgürtel der Erde. Da durch die Lorentz-Kraft die von Sonne und Weltall einfallenden geladenen Teilchen abgelenkt werden, ist es auch eine Vorbedingung für die Erde als Lebensraum.

2.4.2.3. Das Magnetfeld des Leitungsstromes. So wie eine Magnetnadel durch ihr Verhalten das Magnetfeld der Erde anzeigt, zeigt sie in der Nähe eines gleichstromführenden Leiters dessen Magnetfeld an. Jeder stromführende Leiter ist von einem Magnetfeld umgeben, dessen Kraftlinien ihn in Ebenen senkrecht zur Stromrichtung – vom Plus- zum Minuspol zeigend – wie konzentrische Kreise im Rechtsschraubensinn umschließen (Abb. 2.4.2–4).

Der Wert von H ist proportional zur Stromstärke I und umgekehrt proportional zum Abstand r vom Draht $H = I/(2\pi r)$. Die Einheit der magnetischen Feldstärke H ist A/m. Allgemeiner bringt die erste *Maxwellsche Gleichung* $\oint \boldsymbol{H}\cdot d\boldsymbol{s} = I$ zum Ausdruck, daß die Ströme die *Wirbel* des Magnetfeldes sind. Die differentielle Form ist rot $\boldsymbol{H} = \boldsymbol{j}$; sie enthält die

elektrische Stromdichte j. Zusammen mit der Aussage, daß die magnetischen Dipole *Quellen* von $\boldsymbol{H}$ sind, bildet obige Aussage die Grundgleichung der Magnetostatik.

Das *Biot-Savartsche Gesetz* $\mathrm{d}\boldsymbol{H} = I\,\mathrm{d}\boldsymbol{s} \times (\boldsymbol{r}-\boldsymbol{s})/4\pi|\boldsymbol{r}-\boldsymbol{s}|^3$ (Abb. 2.4.2–5) beschreibt den Beitrag eines Linienelementes d$\boldsymbol{s}$ eines stromführenden Drahtes zum Magnetfeld im Raumpunkt P am Ort $\boldsymbol{r}$. Wird der stromführende Leiter zu einer Leiterschleife geschlossen, dann umschließt er eine Fläche, durch die die Kraftlinien des von ihm erzeugten Magnetfeldes hindurchtreten. Die Fläche stellt ein *magnetisches Blatt* dar. Es hat wie jede Magnetnadel und wie jeder Stabmagnet ein *magnetisches Moment* $\boldsymbol{m}$. Die Seite, aus der die Kraftlinien heraustreten, entspricht dem Nordpol einer Magnetnadel, die andere dem Südpol. $\boldsymbol{m}$ ist bei einem ebenen magnetischen Blatt gleich dem Produkt aus Stromstärke und Fläche $I\,A$.

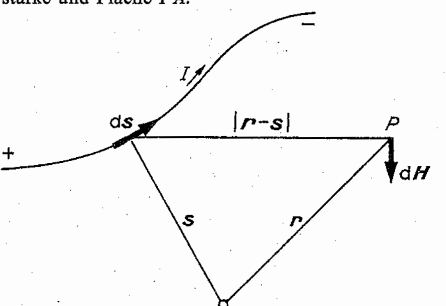


Abb. 2.4.2–5 Beitrag d$\boldsymbol{H}$ des stromführenden Leiterelementes d$\boldsymbol{s}$ zur magnetischen Feldstärke im Raumpunkt P

Eine von einem Gleichstrom I durchflossene Spule erzeugt in ihrem Außenraum ein magnetisches Feld (Abb. 2.4.2–6), das dem eines Stabmagneten vergleichbar ist. Die magnetische Feldstärke $\boldsymbol{H}$ im Innenraum einer im Verhältnis zu ihrem Durchmesser sehr langen Spule ist homogen und dem Betrage nach gleich dem Produkt aus dem Spulenstrom und der auf eine Längeneinheit bezogenen Anzahl von Spulenwindungen, $H = In/l$.

2.4.2.4. Das Magnetfeld des Verschiebungsstromes. Wird ein Kondensator aufgeladen oder entladen, fließt in den Zuleitungen ein Strom. Die Stromstärke I bewirkt auch die zeitliche Änderung der Ladung auf den Metallflächen des Kondensators $\mathrm{d}Q/\mathrm{d}t$ und damit gleichzeitig eine zeitliche Änderung der elektrischen Flußdichte D bzw. der elektrischen Feldstärke E zwischen den Kondensatorplatten, $I = \mathrm{d}Q/\mathrm{d}t = A\,\mathrm{d}D/\mathrm{d}t = A\,\varepsilon\mathrm{d}E/\mathrm{d}t = d\Psi_e/\mathrm{d}t$.

Sowohl der Leitungsstrom I als auch der von Maxwell eingeführte Verschiebungsstrom $\mathrm{d}\Psi_e/\mathrm{d}t$ sind in gleicher Weise mit einem Magnetfeld verknüpft, so daß die erste *Maxwellsche Gleichung* unter Einbeziehung des Verschiebungsstromes $\oint \boldsymbol{H}\mathrm{d}\boldsymbol{s} = I + \mathrm{d}\Psi_e/\mathrm{d}t$ lautet. Die differentielle Form $\mathrm{rot}\,\boldsymbol{H} = \boldsymbol{j} + \partial\boldsymbol{D}/\partial t$ enthält neben der Stromdichte j die zeitliche Änderung der dielektrischen Verschiebung $\boldsymbol{D}$.

2.4.2.5. Magnetische Eigenschaften der Stoffe. Das Magnetfeld $\boldsymbol{H}$ in einer Ringspule wird durch den es erregenden Spulenstrom bestimmt. Es ist unabhängig von dem in den Spulenraum eingebrachten Stoff. Daher wird es auch die magnetische Erregung genannt. Die Kraftflußdichte oder *magnetische Induktion* $\boldsymbol{B}$ dagegen hängt von der Permeabilität μ des eingebrachten Stoffes ab, $\boldsymbol{B} = \mu\boldsymbol{H}$. Für das Vakuum gilt $\boldsymbol{B} = \mu_0\boldsymbol{H}$. Der Zahlenfaktor, mit dem die *magnetische Feldkonstante* μ_0 (vgl. 2.4.3.1) zu multiplizieren ist, um die Permeabilität μ zu erhalten, charakterisiert als relative Permeabilität μ_r die magnetische Eigenschaft des Stoffes: $\mu_r < 1$ diamagnetisch; $\mu_r > 1$ paramagnetisch; $\mu_r \gg 1$ ferromagnetisch. Die Kraftflußdichte denkt man sich zusammengesetzt aus dem Anteil $\mu_0\boldsymbol{H}$, der auch im Vakuum immer vorhanden ist, und aus der Magnetisierung $\boldsymbol{M}$ des den Raum erfüllenden Stoffes $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$. Da es keine einzelnen magnetischen Pole gibt, verschwindet stets der magnetische Fluß aus einer geschlossenen Fläche $\oiint \boldsymbol{B}\cdot\mathrm{d}\boldsymbol{A} = 0$, was in differentieller Form $\mathrm{div}\,\boldsymbol{B} = 0$ bedeutet. Das $\boldsymbol{B}$-Feld hat nur geschlossene Feldlinien (vgl. Abb. 2.4.2–6b). Es ist *quellenfrei.*

Die Magnetisierung ist das Analogon zur elektrischen Polarisation $\boldsymbol{M} = \chi_m\boldsymbol{H}$ mit $\chi_m = \varepsilon_r - 1$ (vgl. 2.4.1.5.) als der *magnetischen Suszeptibilität.* Dia-, para- und ferromagnetische Stoffe haben qualitativ unterschiedliche atomare magnetische Eigenschaften (vgl. 6.7.5.). Für dia- und paramagnetische Stoffe kann die Permeabilität als eine vom Magnetfeld unabhängige Stoffkonstante angesehen werden, bei ferromagnetischen Stoffen ist sie aber eine Funktion der erregenden Feldstärke $\boldsymbol{H}$, d. h., der Zusammenhang zwischen Magnetisierung und Magnetfeld ist nichtlinear. χ_m ist für diamagnetische Stoffe temperaturunabhängig. Für paramagnetische Stoffe gilt das *Curie-Weißsche Gesetz* $\chi_m = C/(T - T_C)$.

Diamagnetismus. Feste längliche diamagnetische Körper stellen sich frei drehbar aufgehängt quer zum magnetischen Feld zwischen den Polschuhen eines Elektromagneten. Sie verhalten sich analog den dielektrischen Körpern in einem Medium mit größerer Dielektrizitätskonstanten (vgl. 2.4.1.5.). Die in ihnen erzeugte Magnetisierung ist dem magnetisierenden Feld entgegengerichtet.

Paramagnetismus ist ein Analogfall zur elektrischen Orientierungspolarisation (vgl. 2.4.1.5.) und bewirkt die Verstärkung der magnetischen Kraftflußdichte aufgrund

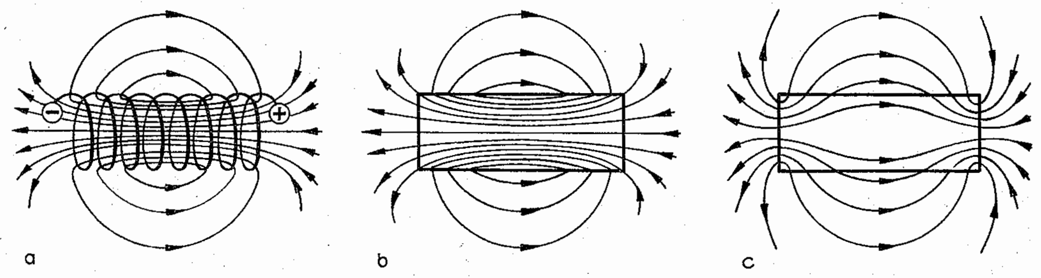


Abb. 2.4.2–6 (a) Magnetfeld $\boldsymbol{B} = \mu_0\boldsymbol{H}$ einer stromdurchflossenen Spule, (b) $\boldsymbol{B}$-Feld und (c) $\boldsymbol{H}$-Feld eines Stabmagneten, $\boldsymbol{B} = \mu_0(\boldsymbol{H} + \boldsymbol{M})$

der Orientierung bereits in den Atomen vorhandener magnetischer Momente in Richtung des magnetisierenden Feldes $\boldsymbol{H}$. Er überwiegt meist den Diamagnetismus weitgehend. Ein einfaches Verfahren zur Messung der Permeabilität einer paramagnetischen Flüssigkeit beruht darauf, daß diese in einem U-Rohrschenkel zwischen den Polen eines Magneten aufsteigt (Abb. 2.4.2–7). Ihre Steighöhe gegenüber dem Flüssigkeitsspiegel in dem anderen U-Rohrschenkel ist $h = \boldsymbol{BH}(\mu_r - 1)/(2\varrho g)$.

Ferromagnetismus setzt den kristallinen Zustand des Stoffes voraus. Er unterscheidet sich dadurch vom Paramagnetismus, daß in mikroskopisch kleinen Kristallbereichen die magnetischen Momente der Atome durch spontane Magnetisierung gleichgerichtet sind. Oberhalb der Curie-Temperatur, auch *Curie-Punkt* genannt, wird, durch die Wärmebewegung der Atome verursacht, ein ferromagnetischer Stoff paramagnetisch.

Bei der Magnetisierung eines ferromagnetischen Stoffes kommt es immer zu einer Sättigungserscheinung (Abb. 2.4.2–8). Bei entsprechend großer Feldstärke H nimmt die Magnetisierung M nur noch sehr wenig zu. Bei Zurücknahme von H auf Null bleibt eine Restmagnetisierung M_R, die erst durch ein Magnetfeld in entgegengesetzter Richtung (Koerzitivkraft H_C) aufgehoben werden kann. Der ferromagnetische Stoff zeigt beim Ummagnetisieren eine Hysterese. Der von der Hystereseschleife eingeschlossene Flächeninhalt liefert die bei einem Ummagnetisierungszyklus in Wärme umgesetzte Energie je Volumeneinheit $W_m/V = \oint H \mathrm{d}B$.

Antiferromagnetismus, Ferrimagnetismus. Es gibt Stoffe, die sich diamagnetisch verhalten, oberhalb einer bestimmten Temperatur aber paramagnetisch werden. Solange der Stoff ein diamagnetisches Verhalten zeigt, haben benachbarte Atome antiparallel gestellte magnetische Momente, so daß die spontane Magnetisierung nach außen nicht in Erscheinung tritt. Typische Vertreter solcher Antiferromagnetika sind α-Fe_2O_3, MnO, MnF_2. Die der Curie-Temperatur entsprechende Temperatur wird bei diesen Stoffen Néel-Temperatur T_N genannt. Die magnetische Suszeptibilität hat bei T_N ihr Maximum und zeigt oberhalb von T_N eine Temperaturabhängigkeit wie bei Paramagneten. Es gibt weiterhin Stoffe, wie den Magnetit $Fe^{2+}Fe_2^{3+}O_4^{8-}$, in denen zwei Kationenteilgitter eine antiparallele Orientierung der magnetischen Momente haben, in denen sich aber wegen der unterschiedlichen Anzahl der Teilgitterplätze die magnetischen Momente nicht kompensieren. Daher zeigen diese Stoffe nach außen hin ein ferromagnetisches Verhalten. Diese Erscheinung wird Ferrimagnetismus genannt. Wichtige Vertreter der Ferrimagnetika sind Eisen-, Nickel- und Manganoxide, die alle nichtleitend sind. Zu feinem Pulver vermahlen, in gewünschte Formen gepreßt und gesintert ergeben sie – als Ferrite bezeichnet – eine Keramik, die ferromagnetisches Verhalten, wenn auch bei viel geringerer Sättigungsmagnetisierung im Vergleich zu den Ferromagnetika, zeigt. Magnetisch weiche Ferrite werden als HF-Spulenkerne, magnetisch harte Ferrite als Dauermagnete eingesetzt.

2.4.3. Der stationäre elektrische Strom

2.4.3.1. Das Ampere als Basiseinheit der elektrischen Stromstärke. Man betrachtet zwei im Vakuum parallel im Abstand von 1 m angeordnete, geradlinige, unendlich lange Leiter von vernachlässigbar kleinem Querschnitt. Der Strom in diesen beiden Leitern soll dann die Stärke von 1 A haben, wenn er zwischen diesen Leitern je 1 m Länge elektrodynamisch die Kraft $2 \cdot 10^{-7}$ N hervorruft.

Die Kraft auf ein Stück Draht der Länge l ist proportional zu l und zum ihn durchfließenden Strom I_1. Weiterhin ist sie proportional zu der den zweiten Draht durchfließenden Stromstärke I_2 und umgekehrt proportional zum Abstand r zwischen den Drähten, d. h. proportional zum vom Draht 2 am Ort des Drahtes 1 erzeugten Magnetfeld $H_2 = I_2/(2\pi r)$. Man beobachtet also $F = \mu_0 I_1 l H_2 = \mu_0 I_1 I_2 l/(2\pi r)$. Mit der Definition des Ampere wird auch der Wert des Proportionalitätsfaktors μ_0, der *Permeabilität des Vakuums* oder *magnetischen Feldkonstanten*, festgelegt, $\mu_0 = 4\pi \cdot 10^{-7}$ N/A².

Während Meter, Sekunde und Kilogramm unabhängig voneinander definiert sind, ist das Ampere, obwohl es eine Basiseinheit ist, durch eine Meßvorschrift an diese Basiseinheiten gebunden. Die Unsicherheit, mit der das Ampere experimentell dargestellt werden kann, beträgt dadurch einige 10^{-6}. Sie überträgt sich auf die abgeleitete Einheit 1 V, die als Spannungsabfall an der Widerstandseinheit 1 Ω darstellbar wird, wenn ein Strom von 1 A fließt. Dabei kann vorausgesetzt werden, daß sich das Ohm mit einer weitaus besseren Genauigkeit darstellen läßt sich das Ampere. So

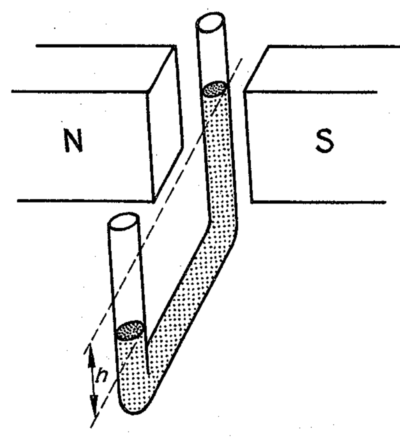


Abb. 2.4.2–7 Paramagnetische Flüssigkeit im Magnetfeld zwischen den Polschuhen eines Magneten

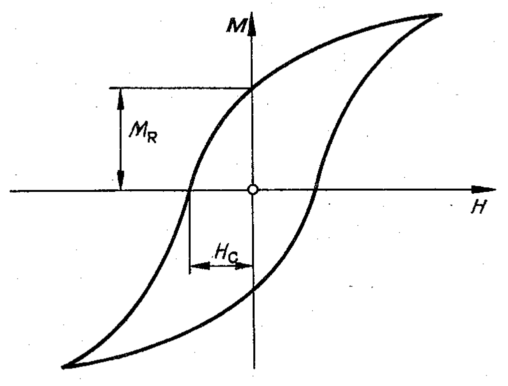


Abb. 2.4.2–8 Sättigung der Magnetisierung und Hysteresekurve eines ferromagnetischen Stoffes

z. B. als *Blindwiderstand* eines Kondensators, dessen Kapazität aus seinen geometrischen Abmessungen und der Feldkonstante ε_0 berechnet, mit einer genau bekannten Frequenz betrieben und in einer Brückenschaltung auf einen Gleichstromwiderstand übertragen wird.

Der *Josephson-Effekt* (vgl. 8.1.4.4.) ermöglicht es, die Basiseinheit Ampere von den mechanischen Einheiten Meter und Kilogramm zu lösen und auf Naturkonstanten zu beziehen. An zwei stromdurchflossenen Supraleitern, die durch eine dünne isolierende Schicht getrennt sind, tritt bei gleichzeitiger Einstrahlung einer Mikrowelle eine Gleichspannung $U = hf/2e$ auf (h = Plancksches Wirkungsquantum, f = Frequenz der Mikrowellen, e = Elementarladung).

2.4.3.2. Gleichstromkreis, Gleichspannungsquellen. Ein Gleichstromkreis besteht aus einer Gleichspannungsquelle und einem Gleichstromverbraucher. Innenwiderstand R_i der Spannungsquelle und Widerstand R des Verbrauchers begrenzen die Stromstärke I. Die elektromotorische Kraft E des Kreises wird nur von der Spannungsquelle aufgebracht, es gilt das *Ohmsche Gesetz* $I = E/(R_i + R)$. Da ein Teil der Spannung am Innenwiderstand der Spannungsquelle abfällt, ist die dem Verbraucher zur Verfügung stehende *Klemmenspannung* U um den Betrag IR_i kleiner als die elektromotorische Kraft, $U = E - IR_i$. Der Innenwiderstand der Spannungsquelle begrenzt auch ihren *Kurzschlußstrom*, $I_K = E/R_i$. Bei einem räumlich ausgedehnten Verbraucher wird die Stromdichte j in jedem Raumpunkt von dem Potentialgefälle bestimmt, $j = -\sigma \operatorname{grad} \varphi$ (vgl. 2.4.1.1.). σ ist die *elektrische Leitfähigkeit*, eine Materialkonstante. Das Ohmsche Gesetz läßt sich so auch als eine Materialgleichung formulieren, $j = -\sigma E$. In einem homogenen Material ist der Widerstand der Länge des Stromweges direkt und dem zur Verfügung stehenden Querschnitt umgekehrt proportional, $R = \varrho l/A$. $\varrho = 1/\sigma$ ist der *spezifische Widerstand* des Materiales. Man beachte, daß die Buchstaben σ und ϱ auch für die Flächen- bzw. Raumladungsdichte verwendet werden (vgl. 2.4.1.1.)! Die Einheit des Widerstandes ist Ohm (Ω), ihr Kehrwert Siemens (S) ist die Einheit des Leitwertes 1 S ($1\,\mathrm{S} = 1\,\Omega^{-1}$).

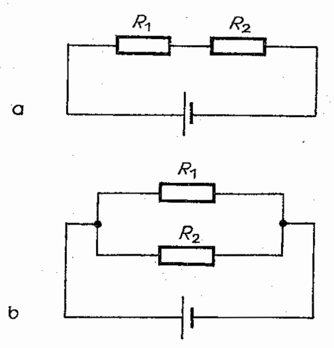
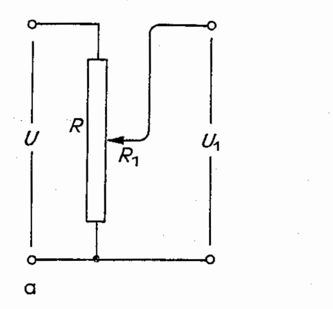


Abb. 2.4.3–1 (a) Reihenschaltung und (b) Parallelschaltung von Widerständen



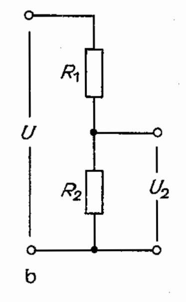


Abb. 2.4.3–2 Spannungsteiler (a) als Potentiometer ($U_1/U = R_1/R$), (b) mit Festwiderständen ($U_2/U = R_2/(R_1 + R_2)$)

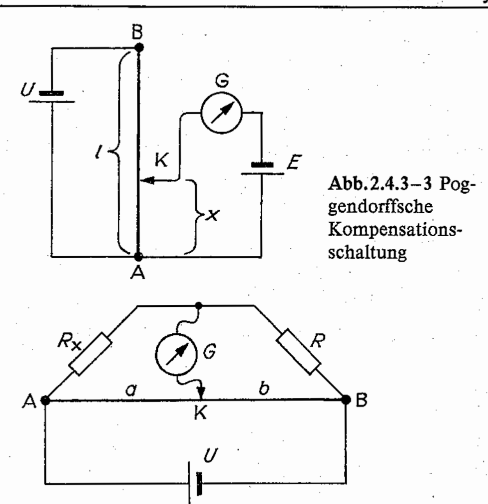


Abb. 2.4.3–3 Poggendorffsche Kompensationsschaltung

Abb. 2.4.3–4 Wheatstonesche Brückenschaltung

Fließt im Gleichstromkreis die Ladung $\mathrm{d}Q = I\mathrm{d}t$ durch den Verbraucher, so durchläuft sie dabei eine Potentialdifferenz $\varphi_2 - \varphi_1 = U$, die durch die Klemmenspannung U gegeben ist. Die Ladung $\mathrm{d}Q$ verringert dabei ihre elektrische Energie um $U\mathrm{d}Q$, so daß die je Zeiteinheit freiwerdende Energie $\mathrm{d}W/\mathrm{d}t = U\mathrm{d}Q/\mathrm{d}t = U\,I = I^2R$ ist. Sie ist gleich dem Produkt aus Spannung und Stromstärke. Ihre Einheit ist Watt ($1\,\mathrm{W} = 1\,\mathrm{V} \cdot 1\,\mathrm{A} = 1\,\mathrm{J/s} = 1\,\mathrm{Nm/s}$). Diese Leistung muß von der Gleichspannungsquelle zeitkonstant bereitgestellt werden, wenn ein Gleichstrom fließen soll.

Ist der Verbraucher ein Widerstandsnetzwerk, kommt es an den Leitungsknoten zur Stromverzweigung. Unterscheidet man in bezug auf den Knoten zu- und abfließende Ströme durch das Vorzeichen, so folgt aus dem Satz von der Erhaltung der Ladung $\sum_i J_i = 0$, für die Hintereinanderschaltung von Widerständen $R = R_1 + R_2$ und für die Parallelschaltung von Widerständen $1/R = 1/R_1 + 1/R_2$. Entsprechend folgt aus dem Satz von der Erhaltung der Energie, daß in jeder Masche eines passiven Widerstandsnetzwerkes $\sum_i R_i I_i = 0$ gelten muß, wenn man die Vorzeichen der Ströme auf ein und denselben Umlaufsinn bezieht (Abb. 2.4.3–1). Enthält eine Masche aktive Spannungsquellen U_j, dann gilt $\sum_i R_i I_i = \sum_j U_j$. Die Berechnung des Gesamtwiderstandes eines Netzwerkes erfordert gegebenenfalls die Umrechnung einer Widerstandsmasche in eine Widerstandsverzweigung.

Spannungsteiler oder *Potentiometer* sind Widerstände, an denen man eine Teilspannung der an ihnen anliegenden Spannung abgreifen kann (Abb. 2.4.3–2). Die *Poggendorffsche Kompensationsschaltung* ist eine Potentiometerschaltung (Abb. 2.4.3–3), bei der mit dem Schleifkontakt K am

Widerstandsdraht AB der Länge l eine der elektromotorischen Kraft E gleiche Spannung abgegriffen wird. Zur Kontrolle der hierfür einzuhaltenden Stromlosigkeit des Abgriffes dient das Nullstrominstrument G. Es gilt $x/l = E/U$.

Die *Wheatstonesche Brücke* dient zur Messung von Widerständen durch Vergleich (Abb. 2.4.3–4). Parallel zum Widerstandsdraht AB, an dem die Klemmenspannung U eines Akkus linear abfällt, sind ein bekannter und ein zu messender Widerstand R und R_x geschaltet. Zwischen beiden und dem Schleifkontakt K liegt die „Brücke" mit dem Nullstrominstrument G. Bei Stromlosigkeit der Brücke gilt $a/b = R_x/R$.

Gleichspannungsquellen. Alle von selbst in einer Anordnung sich einstellenden elektrischen Potentialdifferenzen sind Gleichspannungsquellen. Die Ladungstrennung beruht auf atomaren Vorgängen, durch die die Größe der Potentialdifferenz festgelegt ist. Die auf dem Dynamoprinzip beruhenden Gleichstrommaschinen sind mit Ausnahme der Unipolarmaschinen (vgl. 2.4.4.1.) keine Gleichspannungsquellen im eigentlichen Sinne, da sie die Potentialdifferenz bzw. die elektromotorische Kraft nicht als eine von der Natur der Anordnung fest vorgegebene Größe herstellen, sondern durch einen sich periodisch wiederholenden Vorgang.

Die von *galvanischen Elementen* aufgebrachte Spannung beruht darauf, daß ein mit einem Elektrolyten in Kontakt gebrachtes Metall als Elektrode eine *Lösungstension* aufweist. Es gehen positiv geladene Metallionen in Lösung und bilden mit den Elektronen an der Oberfläche der Elektrode eine elektrische Doppelschicht. Zwischen der Lösung und dem Inneren des Metalles besteht eine Spannung, die für die Kombination Metall-Elektrolyt charakteristisch ist. Zwei gleiche Metallelektroden in die Lösung eines ihrer Metallsalze getaucht, bilden bei unterschiedlicher Konzentration der Lösung eine *Konzentrationskette* (Abb. 2.4.3–5). Die Metallelektrode, die in die Lösung mit geringerer Konzentration eintaucht, ist gegenüber der anderen negativ geladen. Die Spannung zwischen beiden Metallelektroden ist $U = RT2\bar{u}\ln(c_1/c_2)/[Fz(\bar{u}+\bar{v})]$ (R = Gaskonstante, T = Temperatur, F = Faraday-Konstante, z = Wertigkeit des Metallions in Lösung, c_1 und c_2 = Konzentration der Metallsalzlösungen, $\bar{u}$ und $\bar{v}$ = Beweglichkeit des Metallions bzw. des Metallsalzanions.

In einem über einen Verbraucher R kurzgeschlossenen galvanischen Element, z. B. Cu-Zn-Element (Daniell-Element), geht dauernd Zn^{++} in Lösung, Cu^{++} scheidet sich an der Cu-Elektrode ab (Abb. 2.4.3–6). Über den Verbraucher fließen Elektronen vom Zink zum Kupfer und neutralisieren die dort im Elektrolyten ankommenden Cu^{++}-Ionen. Der elektrische Strom ist – wegen der negativen Ladung der Elektronen – im Verbraucher also von der Cu- zur Zn-Elektrode gerichtet. Die Cu-Elektrode ist die positive Elektrode. Bei einem galvanischen Element Platin-Kupfer wäre die Cu-Elektrode die negative. Platin, Kupfer und Zink sind nach steigender Lösungstension geordnet. An der Phasengrenze zwischen Elektrode und Elektrolyt ist die Lösungstension als Potentialdifferenz nicht meßbar. Um trotzdem die verschiedenen Elektrodenmaterialien in einer *elektrochemischen* oder *Voltaschen Spannungsreihe* ordnen zu können, gibt man ihr Potential in bezug auf eine Normalwasserstoffelektrode an, deren Potential willkürlich gleich Null gesetzt ist (Tab. 2.4.3–7).

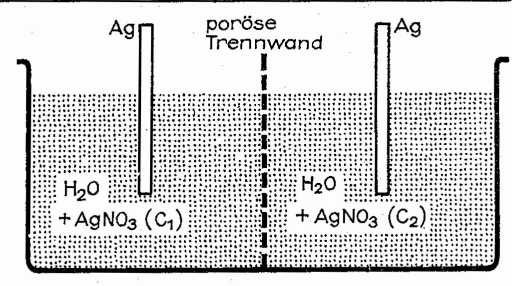


Abb. 2.4.3–5 Konzentrationskette mit poröser Trennwand

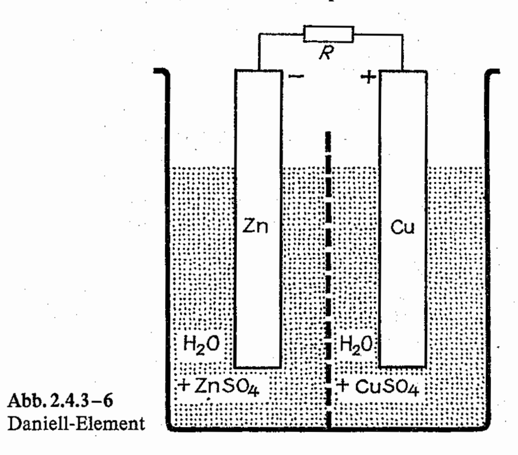


Abb. 2.4.3–6 Daniell-Element

Tab 2.4.3–7 Elektrochemische Spannungsreihe der Metalle (Normalspannung gegen Wasserstoffelektrode in Volt)

Li^+	−3,02	Co^{2+}	−0,28
K^+	−2,92	Ni^{2+}	−0,24
Ca^{2+}	−2,76	Sn^{2+}	−0,16
Na^+	−2,71	Pb^{2+}	−0,126
Mg^{2+}	−2,34	$2\,H^+$	±0,00
Al^{3+}	−1,33	Cu^{2+}	+0,345
Mn^{2+}	−1,10	Ag^+	+0,79
Zn^{2+}	−0,762	Hg^{2+}	+0,85
Cr^{3+}	−0,51	Au^{3+}	+1,36
Fe^{2+}	−0,44	Pt^{2+}	+1,60
Cd^{2+}	−0,402		

Normalelemente dienen als Eichelemente für die Spannungsmessung. Das stromlose *Weston-Element* hat bei 20 °C eine Spannung von 1,01865 V. Die Strombelastung darf 0,1 A nicht überschreiten.

Thermoelemente sind Gleichspannungsquellen, deren Potentialdifferenz als *Kontakt-* oder Berührungsspannung dadurch zustande kommt, daß bei den zwei verschiedenen in Kontakt befindlichen Metallen die Leitungselektronen zu dem Metall übertreten, dessen Fermi-Energie niedriger bzw. dessen Austrittsarbeit für Elektronen größer als die des anderen ist.

Ein zu einem Leiterkreis geschlossenes Thermoelement ist stromlos, wenn sich beide Kontaktstellen auf der gleichen Temperatur befinden. Die Kontaktspannungen sind dem Betrage nach gleich und entgegengerichtet. Sind die beiden Kontaktstellen auf verschiedenen Temperaturen, resultiert wegen der Temperaturabhängigkeit der Kontaktspannung eine sogenannte Thermospannung, es fließt ein Strom (*See-*

Tab. 2.4.3–8 Thermoelektrische Spannungsreihe der Metalle (Thermokraft gegen Kupfer in μV K⁻¹)

Se	+997	Au	+0,1	Pt	−5,9
Te	+397	Cu	±0,0	Hg	−6,0
Ge	+297	Re	−0,1	Na	−7,0
Sb	+32	Ag	−0,2	Pd	−8,3
Fe	+13,4	W	−1,1	Ca	−10,8
Li	+8,7	Cs	−2,4	K	−14,3
Ce	+4,4	Sn	−2,6	Co	−20,1
Mo	+3,2	Pb	−2,8	Ni	−20,4
Zn	+0,3	Mg	−3,0	Bi	−72,8
Cd	+0,2	Al	−3,2		

beck-Effekt). Die auf eine Temperaturdifferenz von 1 K bezogene Thermospannung wird als Thermokraft bezeichnet. Sie ist charakteristisch für die jeweilige Metallkombination und i. allg. eine Funktion der Temperatur. In der *thermoelektrischen Spannungsreihe* sind die Metalle nach ihrer Thermokraft, bezogen auf ein bestimmtes Metall, geordnet (Tab. 2.4.3–8). Fließt durch eine Kontaktstelle von zwei Metallen ein Strom so, wie er als Thermostrom durch die kältere Kontaktstelle fließen würde, wird diese erwärmt, andernfalls abgekühlt *(Peltier-Effekt)*.

2.4.3.3. Elektrizitätsleitung in metallischen Leitern, Halbleitern, Elektrolyten und Gasen. Seit dem Nachweis der Elektronen im Atom und der Katodenstrahlteilchen als freie Elektronen Ende des vorigen Jahrhunderts nahm man an, daß in den Metallen die Elektrizitätsleitung durch Elektronen erfolgt. Es gab dafür zwei wichtige Hinweise: Beim Stromfluß durch Metalle kommt es nicht zu einem Materietransport. Das Reflexionsvermögen der Metalle ist im gesamten Spektrum der elektromagnetischen Wellen vom UV-Licht bis zu den Radiowellen durchgängig sehr hoch. Der direkte Nachweis quasifreier Elektronen als Stromträger in Metallen wurde durch die Versuche von Tolman in den 20er Jahren erbracht. Seitdem zählt man feste und flüssige Leiter dann zu den Metallen, wenn ihre Elektrizitätsleitung überwiegend durch Elektronen erfolgt und die Anzahl der für die Elektrizitätsleitung zur Verfügung stehenden Elektronen von der gleichen Größenordnung ist, wie die Anzahl der im Leiter vorhandenen Atome.

Ionenleitung tritt in Metallen nur bei Verunreinigung mit nichtmetallischen Elementen auf. So wandert z. B. der Kohlenstoff in einem glühenden Eisendraht zur Katode.

Die Driftgeschwindigkeit der Elektronen, wie überhaupt aller Ladungsträger in Leitern, ist der elektrischen Feldstärke proportional, $\bar{u} = bE$ (b = Ladungsträgerbeweglichkeit). In der Zeiteinheit passieren $nA\bar{u}$ Ladungsträger den Leiterquerschnitt A, die elektrische Leitfähigkeit wird darstellbar durch die räumliche Dichte n der Ladungsträger mit der Elementarladung e und der Beweglichkeit b, $j = \sigma E = enbE$.

Der spezifische elektrische Widerstand $\varrho = 1/\sigma$ der metallischen Leiter nimmt mit wachsender Temperatur zu. Sein Temperaturkoeffizient $\alpha = (1/\varrho)\, \mathrm{d}\varrho/\mathrm{d}T > 0$ kommt für Platin dem Wert 1/273 je Kelvin, dem Ausdehungskoeffizienten der Gase, sehr nahe. Für die anderen Metalle liegt er nur wenig höher. Beim flüssigen Quecksilber liegt er bei

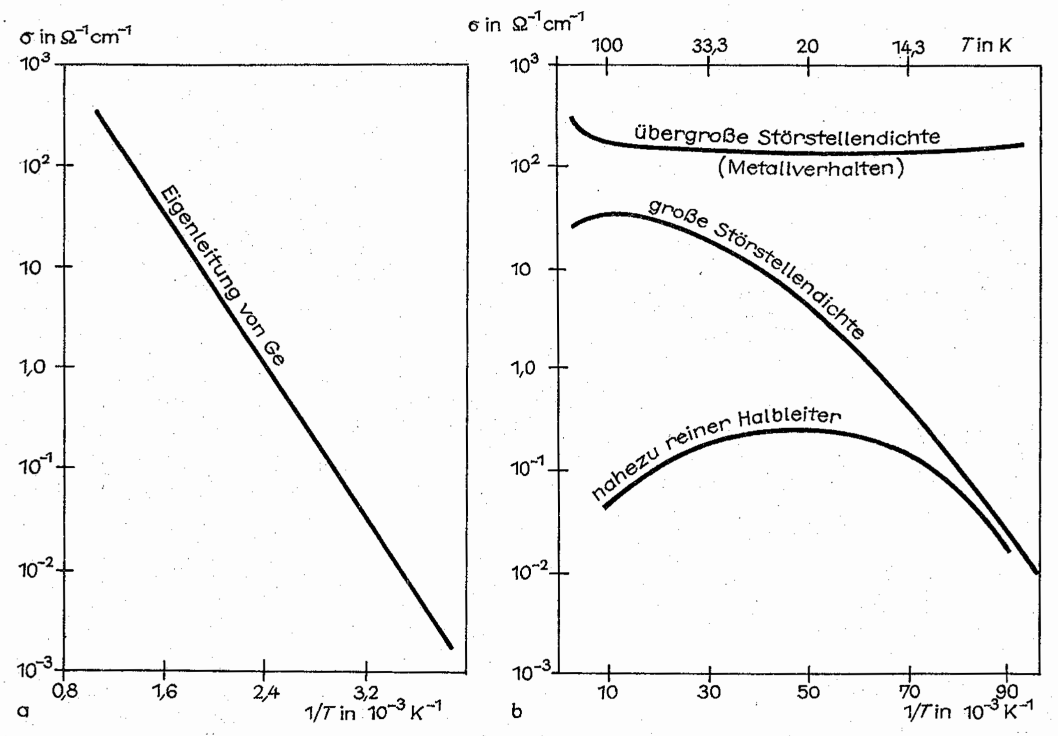


Abb. 2.4.3–9 (a) elektrische Leitfähigkeit von Germanium Ge bei Eigenleitung, (b) elektrische Leitfähigkeit von n-leitendem Ge als Funktion der Temperatur (nach Madelung)

10^{-3} je Kelvin. Gehen die Metalle in den Dampfzustand über, verlieren sie ganz ihre metallische Eigenschaft, Elektronenleiter zu sein.
Alle Störungen des Kristallgitters der Metalle ändern den Einfluß der Temperatur auf den spezifischen Widerstand, der sonst oberhalb 20 K der Temperatur nahezu proportional ist. Es gilt die empirisch gefundene *Matthiessensche Regel*, nach der das Verhältnis aus Temperaturkoeffizient α und der elektrischen Leitfähigkeit σ bei einer gegebenen Temperatur konstant ist, unabhängig davon, welche Gitterstörung das Metall durch Verunreinigung oder durch Verformung erfahren hat, $\alpha/\sigma =$ const. Bei Abkühlung bis in die Nähe des absoluten Nullpunktes weisen alle Metalle – die supraleitenden können durch ein überkritisches Magnetfeld zu Normalleitern werden – einen *Restwiderstand* auf, der ein summarisches Kriterium für Reinheit und Kristallbaufehlerfreiheit ist.
Im Gegensatz zu den reinen Metallen haben die Metallegierungen i. allg. einen um eine Größenordnung größeren spezifischen Widerstand. Für die Herstellung von Präzisionswiderständen für Meßzwecke sind Legierungen entwickelt worden, deren Temperaturkoeffizient α bis zu drei Größenordnungen kleiner als der von den gebräuchlichen Leitermaterialien ist.
Halbleiter. In der Nähe des absoluten Nullpunktes sind reine Halbleiter Isolatoren. Mit ansteigender aber noch sehr niedriger Temperatur ist eine Elektronenleitfähigkeit nachweisbar. Sie wird durch Gitterstörstellen ermöglicht. Zur Eigenleitung durch Bereitstellung von Ladungsträgern aus dem Gitter durch Ionisierung von Gitteratomen kommt es erst bei höheren Temperaturen. Ein an das Atom gebundenes *Valenzelektron* wird zum Leitungselektron. Ein positiver Atomrumpf bleibt zurück. Er übernimmt von einem katodenseitig gelegenen Atom ein Valenzelektron und gibt so die positive Ladung in Richtung Katode weiter. Dieser Vorgang wiederholt sich, bis die positive Ladung, das *Defektelektron*, die Katode erreicht hat und dort von einem Leitungselektron aus der Zuleitung neutralisiert wird. Bei der Eigenleitung werden somit gleichviele Elektronen wie Defektelektronen „erzeugt“. Es bleibt nicht aus, daß ein gewisser Anteil von Defektelektronen Leitungselektronen einfängt und somit beide für den Ladungstransport wieder ausfallen. Die Temperaturabhängigkeit der elektrischen Leitfähigkeit eines reinen Halbleiters zeigt Abb. 2.4.3–9.
Liegt kein reiner Halbleiter vor, oder ist bei der Herstellung des Halbleiters bewußt eine Dotierung mit Fremdatomen vorgenommen worden, ist ein entsprechender Anteil der Gitterplätze mit Fremdatomen besetzt. Haben sie mehr Valenzelektronen, als sie zur Bindung an die Atome ihrer Umgebung benötigen, lassen sich die überschüssigen mit geringerer Energiezufuhr zu Leitungselektronen machen, der Halbleiter ist n-leitend, im entsprechend umgekehrten Fall p-leitend. Da die Driftgeschwindigkeit von Elektronen und Defektelektronen i. allg. verschieden ist und je nach Dotierung auch ihre Ladungsträgerdichten, geht die für Metalle gültige Beziehung zwischen Stromdichte und elektrischer Feldstärke über in die für isotrope Halbleiter gültige Beziehung $j = e(n_p b_p - n_n b_n)E$, worin die Indizes n und p negativ bzw. positiv bedeuten.
Elektrolyte sind Ionenleiter. Mit dem Ladungstransport erfolgt gleichzeitig ein Stofftransport, eine Elektrolyse.
Feste Ionenleiter, sogenannte Feststoffelektrolyte, sind i. allg. kristalline Verbindungen mit überwiegend heteropolarem Bindungsanteil der Atome bzw. Atomgruppen von zwei ineinandergestellt zu denkenden Teilgittern, von denen das eine fest bleibt, das andere aufgrund, einer Fehlordnung die Platzwechsel der Ionen und damit deren Beweglichkeit für die Stromleitung ermöglicht. In bestimmten Fällen kann sich diese Beweglichkeit auf eine kristallografische Richtung oder Ebene beschränken. Sie kann so groß werden wie in konzentrierten Elektrolytlösungen oder in Salzschmelzen. Dadurch, daß ein Teilgitter fest bleibt, kann ein fester Ionenleiter auch gleichzeitig als Gefäß oder auch als Trennwand zwischen flüssigen Elektroden fungieren. Alle Ionenkristalle haben aufgrund ihrer Entstehung bzw. Herstellung bei einer Temperatur oberhalb 0 K eine Fehlordnung (vgl. 6.6.). Bei Zimmertemperatur und darunter ist aber ihre Ionenleitung sehr klein und meist zu vernachlässigen. Zu einer für eine merkliche Ionenleitung hinreichenden Fehlordnung kommt es erst bei hohen Temperaturen, gegebenenfalls begünstigt durch Übergang in eine Hochtempera-

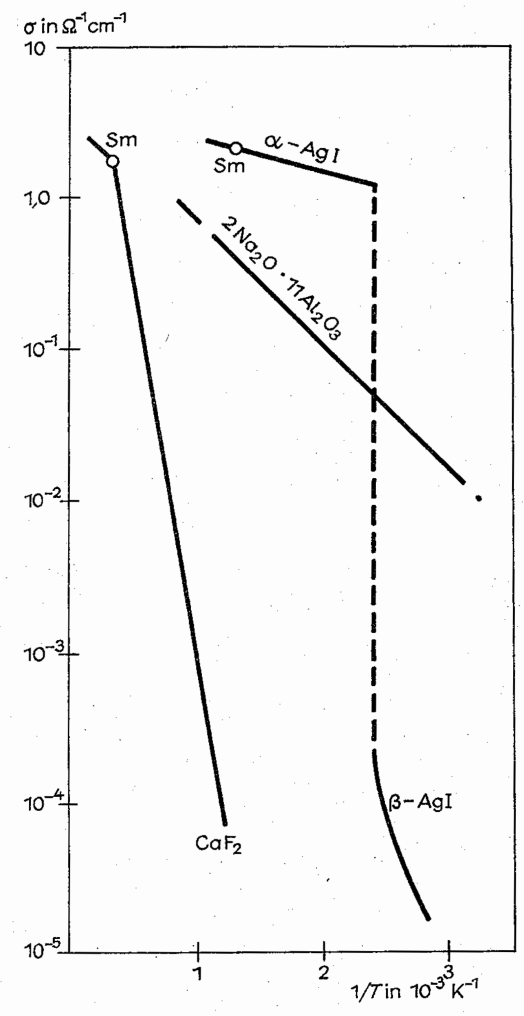


Abb. 2.4.3–10 Elektrische Leitfähigkeit von Feststoffelektrolyten in Abhängigkeit von der Temperatur (nach DROTSCHMANN) (zum Vergleich: $\sigma_{Fe} = 10^5 \Omega^{-1}\text{cm}^{-1}$; $\sigma_{Cu} = 5{,}88 \cdot 10^5 \Omega^{-1}\text{cm}^{-1}$)

turmodifikation oder durch eine Dotierung in der Größenordnung bis zu 10 Molprozent mit Verbindungen von abweichender Wertigkeit der heteropolaren Bindung, so daß das die Ionenleitung übernehmende Teilgitter unterbesetzt ist. Typische Vertreter dieser Fälle sind CaF_2 mit F^--Ionenleitung durch kontinuierlich zunehmende Fehlordnung mit der Temperatur, α-AgI mit Ag^+-Ionenleitung nach Übergang von der hexagonalen β- zur kubischen α-Modifikation und β-$Al_2O_3 = 2Na_2O \cdot 11Al_2O_3$ mit Na^+-Ionenleitung als ein mit Na_2O dotiertes Al_2O_3 (Abb. 2.4.3–10).

Flüssige Ionenleiter sind Salzschmelzen und Elektrolytlösungen. Hierzu zählen auch Säuren und Laugen und ihre wäßrigen Lösungen. Hochkonzentrierte Salzlösungen können als Salzschmelzen aufgefaßt werden, in denen ein geringer Anteil Lösungsmittel gelöst ist. In allen Fällen entstehen die Ladungsträger durch *elektrolytische Dissoziation.* Unter *Dissoziationsgrad* versteht man das Verhältnis der Anzahl dissoziierter Ionenpaare zu ihrer Gesamtzahl. In den wäßrigen Elektrolytlösungen sind die Ionen von einer Koordinationssphäre von Wassermolekülen umgeben, die sich mit ihrem Dipolmoment zum Ion hin orientiert haben und im inhomogenen elektrischen Feld des Ions eine Anziehung erfahren. Die Bindung in der ersten Koordinationssphäre kann bis zu beständigen Aquokomplexverbindungen führen. Der Durchmesser der Solvathüllen ist in nichtwäßrigen Lösungsmitteln gewöhnlich größer als in wäßrigen. Er bestimmt den Reibungswiderstand, wenn sich das solvatisierte bzw. hydratisierte Ion im elektrischen Feld der Elektroden durch die Lösung hindurchbewegt. Die Beweglichkeit der Ionen beträgt deshalb auch nur den etwa 10000sten Teil der Beweglichkeit der Elektronen in Metallen.

Die beiden in den Elektrolyten eintauchenden Elektroden bilden mit diesem ein galvanisches Element, dessen Spannung U_z als *Polarisations-* oder *Zersetzungsspannung* der für den Vorgang der Elektrolyse angelegten Spannung entgegengerichtet ist, $I = (U - U_z)/R_z$ mit R_z als Widerstand der Elektrolysezelle. Sie kommt dadurch zustande, daß die Elektroden mit dem Zersetzungsprodukt des Elektrolyten belegt werden, z. B. bei wäßrigen Elektrolyten die Katode mit H_2 und die Anode mit O_2.

Der Zusammenhang zwischen den bei der Elektrolyse abgeschiedenen Stoffmengen und der hierfür benötigten Elektrizitätsmenge wird durch die *Faradayschen Gesetze* beschrieben: 1. Die abgeschiedene Stoffmenge ist der Stromstärke I und der Zeit t proportional, $m = KIt$. K ist das *elektrochemische Äquivalent.* Es gibt an, wieviel Gramm von 1 As abgeschieden werden, $K_{Ag} = 1{,}118$ mg/As, $K_{Cu} = 0{,}329$ mg/As. Bei 0 °C und 10^5 Pa werden 0,172 cm^3 Knallgas erzeugt. 2. Die durch gleiche Elektrizitätsmengen abgeschiedenen Mengen verschiedener Elektrolyte verhalten sich zueinander wie die *Äquivalentgewichte* der Stoffe.

Unter dem Äquivalentgewicht eines Stoffes versteht man das Verhältnis aus dem Atomgewicht des Stoffes zu seiner Wertigkeit. Zur Abscheidung von 1 Grammäquivalent (val) eines Stoffes benötigt man die Elektrizitätsmenge $9{,}648544 \cdot 10^4$ As. Der Zahlenwert ist von der chemischen Natur des Stoffes unabhängig und heißt *Faradaysche Zahl.*

Gase sind bei Raumtemperatur und tieferen Temperaturen nichtleitend. Die Gasmoleküle (damit sind gewöhnlich auch die Gasatome der Edelgase, der Metalldämpfe u. a. gemeint) sind neutral. Durch die kosmische Strahlung, die Radioaktivität der Gemäuer und der Luft enthält jede Gasmenge eine wenn auch sehr kleine Zahl von Ionen, die fortlaufend rekombinieren. Im Gegensatz zu den Ionen in den Elektrolyten sind die Ionen in Gasen nicht beständig. Die Abnahme der Ladungsträgerkonzentration durch Rekombination ist den Konzentrationen beider Arten von Ladungsträgern proportional, $dn = -\beta n_+ n_- dt$ (β = Rekombinationskoeffizient). Sind die Ladungsträger als Paare erzeugt worden, ist $n_+ = n_- = n$. Ist in einem Gas die Ladungsträgerdichte konstant, herrscht ein stationärer Zustand, die Zahl der je Zeiteinheit erzeugten Ladungsträgerpaare ist gleich der, die rekombiniert sind, $dn/dt_{neu} = dn/dt_{rekomb.} = \beta n^2$, so daß $n = \sqrt{(1/\beta)(dn/dt)_{neu}}$ ist. Die Zeit t, die verstreicht, bis gerade so viele Ionen neu entstanden sind, wie der stationäre Zustand enthält, bezeichnet man als die *mittlere Lebensdauer* der Ionen, $\tau = 1/(\beta n)$.

Der Stromdurchgang durch ionisierte Gase hat eine Ladungsträgerverarmung zur Folge, sofern nicht ständig neue Ladungsträgerpaare durch Ionisierung erzeugt werden. Er wird deshalb Gasentladung genannt. Wenn die Ladungsträgerpaare durch Fremdeinwirkung entstehen, spricht man von einer *unselbständigen Gasentladung.* Die Strom-Spannungs-Charakteristik zeigt eine Sättigung, weil trotz höherer Spannung nicht mehr Ionen je Zeiteinheit zu den Elektroden geholt werden können, als durch die Fremdeinwirkung je Zeiteinheit erzeugt werden. Die Energie zur Ionisierung der Gasmoleküle oder zur Auslösung von Elektronen aus der Katode wird bei der unselbständigen Gasentladung von außen durch Lichtquanten, Röntgenquanten, γ-Quanten, durch Teilchenstrahlung, durch Wärmezufuhr oder auch durch chemische Vorgänge bei höherer Temperatur bereitgestellt.

In der *Lufthülle der Erde* kommt es dadurch, daß die durch Fremdeinwirkung ionisierten Luftschichten selbst Feldwirkungen ausüben und durch Druckgefälle unter dem Einfluß der Lorentz-Kraft bewegt werden, zu sehr komplizierten Erscheinungen der Elektrizitätsleitung. Die Beweglichkeiten der Luftionen sind bei Normaldruck und 0 °C: $b_p = 0{,}000\,137$ (m/s)/(V/m), $b_n = 0{,}000\,190$ (m/s)/(V/m). In allen Luftschichten kommen Ladungsträger mit größerer Masse hinzu, wie Staub, Salz- und Eiskristalle, Nebeltröpfchen usw. Der Durchschnittswert der täglich und über Land und See unterschiedlich schwankenden Feldstärke beträgt 130 V/m. Die elektrische Leitfähigkeit der Luft, die mit zunehmender Höhe zunimmt, liegt bei $2{,}5 \cdot 10^{-14}\,\Omega^{-1}m^{-1}$. In 30 km Höhe wird mit einer rund 5000mal größeren elektrischen Leitfähigkeit als in Erdbodennähe gerechnet. Mit Fragen der Elektrizitätsleitung in den höheren Luftschichten beschäftigt sich die Ionosphärenforschung.

Bei der *selbständigen Gasentladung* wird die Energie zur Erzeugung von Ladungsträgerpaaren durch die Gasentladung selbst bereitgestellt. Je nach den Versuchsbedingungen gibt es verschiedene Formen selbständiger Gasentladung:

Dunkelentladung. Sie ist der Vorläufer einer Glimmentladung, kann aber als unselbständige Gasentladung bei geringer Stromstärke aufrechterhalten werden. Die Stromstärke folgt dem Ohmschen Gesetz. Sie wird auch *Townsend-Entladung* genannt und technisch zur Erzeugung von Ozon und atomarem Wasserstoff genutzt.

Glimmentladung. Sie ist durch ihre Leuchterscheinungen, die in Wissenschaft und Technik vielfältig benutzt werden, bekannt (Abb. 2.4.3–11). Sie wird eingeleitet durch Beschleunigung der immer in jedem Gas vorhandenen Ionen, die beim Aufprall auf die kalte Katode Elektronen auslösen, die nun von sich aus durch Stoßionisation Ladungsträgerpaare erzeugen. Die freiwerdenden Elektronen bilden

eine Elektronenlawine. Ihr Strom wird nach TOWNSEND mit einem festzulegenden Anfangsstrom I_0 zu $I = I_0 e^{\alpha d} / [1 - \gamma(e^{\alpha d} - 1)]$ veranschlagt. Hierin bedeuten α das Ionisierungsvermögen, d die durchlaufene Strecke und γ die Zahl der Sekundärelektronen je auf die Katode auffallendem Ion. Für $\gamma(e^{\alpha d} - 1) \geqq 1$ verliert der Ausdruck für I seinen physikalischen Sinn, und die Gasentladung geht in eine selbständige über. Es bilden sich die für die Glimmentladung charakteristischen Raumladungen aus. In der Glimmhaut vor der Katode regen die ausgelösten Elektronen die Gasmoleküle zum Leuchten an. Im anschließenden Hittorfschen Dunkelraum finden hauptsächlich die Ionisierungen statt, die eine positive Raumladung aufbauen und für den Katodenfall verantwortlich sind. Im Glimmsaum des negativen Glimmlichtes geben die Elektronen ihre restliche kinetische Energie ab und regen zum Leuchten an. Im negativen Glimmlicht und im anschließenden Faradayschen Dunkelraum ist die Feldstärke sehr klein, sie kann sogar negativ werden. Der Stromtransport erfolgt weitgehend durch Diffusion der Ladungsträger. Die Feldstärke steigt erst wieder vor der positiven Säule an, um bis zum Anodenfall konstant zu bleiben. In der positiven Säule sind positive Ionen und Elektronen in gleicher Ladungsträgerdichte vorhanden. Ein solches Gemisch bezeichnet man als ein quasineutrales Plasma. Der Stromtransport wird wegen ihrer größeren Beweglichkeit überwiegend von den Elektronen getragen. Der sich an die positive Säule anschließende Anodenfall ruft in besonderen Fällen, z.B. in stark elektronegativen Gasen ein positives Glimmlicht und einen Anodendunkelraum hervor.

Bogenentladung. Die Glimmentladung geht in eine Bogenentladung über, wenn durch Überhitzung der Katode die Elektronen durch Glühemission austreten. Für eine Quecksilberkatode gilt das nicht. Die große Elektronenstromdichte bei der Bogenentladung kommt womöglich erst in einer über der Quecksilberoberfläche sich ausbildenden Dampfschicht zustande. Man unterscheidet bezogen auf die Stromstärke noch zwischen Bogenentladung und Hochstrombogen. Von einem *Feldbogen* spricht man, wenn bei relativ niedriger Temperatur der Katode eine Bogenentladung durch eine über der Katode befindliche positive Raumladung eine übergroße Feldwirkung im Sinne eines Katodenfalles zustande kommt.

Funkenentladung. Funken sind ihrem Wesen nach rasch verlöschende Bogenentladungen. Ihnen gehen vorbereitende lichtschwache Entladungen voraus. Eine Kugelfunkenstrecke eignet sich zur Spannungsmessung, da die Überschlagsspannung bei gegebenem Gasdruck und gegebener Elektrodenform reproduzierbar ist (Tab. 2.4.3–12). Gleit-

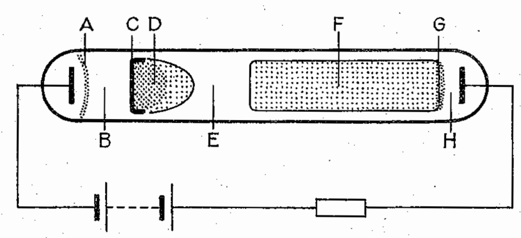


Abb. 2.4.3–11 Leuchterscheinungen der Glimmentladung (nach GRIMSEHL); A = Glimmhaut, B = Hittorfscher Dunkelraum, C = Glimmsaum, D = negatives Glimmlicht, E = Faradayscher Dunkelraum, F = positive Säule, G = Anodenglimmlicht, H = Anodendunkelraum

Tab. 2.4.3–12 Überschlagspannung in Kilovolt einer Kugelfunkenstrecke in trockener Luft von 20 °C bei Normaldruck für zwei gleichgroße Kugeln bei symmetrischer Spannungsverteilung bzw. +) bei Erdung einer Kugel [nach PEEK (1) und ESTORFF (2)]

Schlagweite in cm	Kugeldurchmesser in cm 1,1 (1)	5 (2)	6,6 (1)	7,5 (2)	10 (2)	15 (2)	25 (1)	+)25 (1)
0,05	2,9		2,65				2,65	
0,1	4,95		4,60				4,60	
0,5	17,9		17,5				17,4	
1,0	29,5	32,4	32,2	32,1	32,1	31,8	31,8	31,9
5		111		122	128	133	138	134
10		150		178	199,5	222	247	241
15				208	241	282	331	351
20							398	365
25							451	402
30							492	430
40							555	468

funken bilden sich längs der Oberfläche von Isolatoren aus. Ihnen gehen ebenfalls Vorentladungen voraus. Die Schlagweite von Gleitfunken ist deshalb so groß, weil durch die Vorentladungen an der Oberfläche befindliche Ladungen die Rekombination der Ionen verhindern.

Koronaentladung. Wesentlich für das Zustandekommen der Koronaentladung ist die Krümmung der als Elektrode fungierenden Metallteile. Bei positiver Elektrode werden Gasmoleküle durch das inhomogene Feld vor der Elektrode ionisiert. Bei negativer Metallelektrode ist die zur Entladung erforderliche Spannung um einige 100 V niedriger als bei positiver. Da die Erde gegenüber der Atmosphäre negativ geladen ist, kommt es bei bestimmten Wetterlagen auf natürliche Weise zu Koronaentladungen, zum *Elmsfeuer.*

2.4.3.4. Elektrizitätsleitung im Vakuum. Sie erfolgt durch freie Ladungsträger und wird in Elektronenröhren angewendet. Nach der Anzahl der Elektroden unterscheidet man Diode, Triode usw. Die Elektronen werden durch Glühemission aus der Katode freigesetzt. Ihre kinetische Energie beim Auftreffen auf die Anode ist gleich dem Produkt aus der Elementarladung e und der Anodenspannung U_A. Sie wird – wie auch bei anderen freien Ladungsträgern – in Elektronenvolt angegeben. Sie wird beim Auftreffen auf die Anode in Wärme umgesetzt, die Elektronenröhre ist somit ein Verbraucher im Stromkreis und hat demzufolge einen Widerstand. Bei Anlegen einer Wechselspannung an eine Diode kommt es zu einem Gleichrichtereffekt, da nur dann ein Anodenstrom fließt, wenn die heiße Elektrode Katode ist. Da nicht mehr Elektronen je Zeiteinheit zur Anode abfließen können als die Katode durch Glühemission bereitstellt, kommt es zu einem Sättigungsstrom. Die Sättigungsstromdichte J_s ist abhängig von der Temperatur der Katode und von der Austrittsarbeit W_a der Elektronen aus dem Katodenmaterial, $J_s = aT^2 e^{-W_a/kT}$ mit $a = 60{,}2\,\mathrm{Acm^{-2}K^{-2}}$ *(Richardson-Formel).*

2.4.4. Das zeitlich veränderliche elektromagnetische Feld

2.4.4.1. Die elektromagnetische Induktion. Im elektromagnetischen Feld wirkt auf eine Ladung Q die Lorentz-Kraft $\boldsymbol{F} = Q(\boldsymbol{E} + \boldsymbol{v} \times \boldsymbol{B})$. Sie besteht aus zwei Anteilen, einer Kraftwirkung $Q\boldsymbol{E}$ im elektrischen Feld und einer

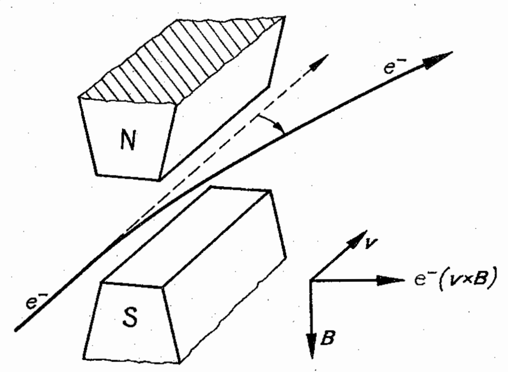


Abb. 2.4.4–1 Ablenkung eines Elektronenstrahles im Magnetfeld

Kraftwirkung $Q\boldsymbol{v}\times\boldsymbol{B}$ im magnetischen Feld, die aber nur auftritt, wenn sich die Ladung im Magnetfeld bewegt. $\boldsymbol{v}$ ist die Geschwindigkeit der Ladung Q relativ zum Magnetfeld $\boldsymbol{B}$. Durch diesen Anteil der Lorentz-Kraft kommt z.B. die Ablenkung eines Elektronenstrahles in einem Magnetfeld zustande (Abb. 2.4.4–1). Weitere Beispiele sind die Ladungstrennung in einem magnetohydrodynamischen Generator (MHD-Generator), die Zurückführung der geladenen Teilchen auf Kreisbahnen in das beschleunigende elektrische Feld zwischen den Duanten eines Zyklotrons (vgl. 5.4.3.1.) und der Hall-Effekt.
Der Anteil $Q\boldsymbol{v}\times\boldsymbol{B}$ stellt aber auch die bei der Induktion ursächlich wirksame elektromotorische Kraft auf die Ladung dar. So bewirkt er beim Bewegen eines Leiters im Magnetfeld ein Verschieben der freibeweglichen Leitungselektronen, während er die im Kristallgitter festsitzenden Ionen nicht von ihren Plätzen wegführt. An den Enden des Leiters tritt eine elektrische Spannung $\int(\boldsymbol{v}\times\boldsymbol{B})\,\mathrm{d}\boldsymbol{s}$ auf, z.B. bei der Ankerwicklung eines Dynamos (Abb. 2.4.4–2b). In einem stromdurchflossenen Leiter, der sich in einem Magnetfeld befindet, bewegen sich die Ladungsträger relativ zum Magnetfeld. Die auf sie einwirkende Lorentz-Kraft überträgt

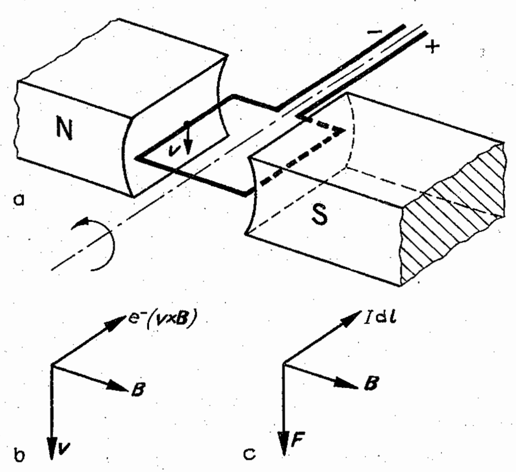


Abb. 2.4.4–2 (a) Leiterschleife im Magnetfeld, (b) Richtungsbeziehungen bei Induktion im Generator, (c) bei Kraftwirkung im Motor

sich auf den Leiter, der dadurch eine mechanisch wirkende Kraft $Q\boldsymbol{v}\times\boldsymbol{B}=I\mathrm{d}\boldsymbol{l}\times\boldsymbol{B}$ erfährt, z. B. die Ankerwicklung eines Motors (Abb. 2.4.4–2c). Darin bedeuten I elektrischer Strom und $\mathrm{d}\boldsymbol{l}$ ein in Stromrichtung weisendes Längenelement des Leiters.
Wird ein Magnet von einer Leiterschleife weg (oder auf sie zu) bewegt (Abb. 2.4.4–3), so beobachtet man ebenfalls eine induzierte Spannung. Diese hat den gleichen Wert wie in dem Fall, in dem sich die Leiterschleife mit derselben Relativgeschwindigkeit von dem Magneten weg bewegt. Daraus läßt sich erkennen, daß eine zeitliche Änderung des Magnetfeldes zu einem elektrischen Feld führt. $\partial\boldsymbol{B}/\partial t$ stellt die *Wirbel des elektrischen Feldes* dar. Der mathematische Ausdruck dafür ist die zweite *Maxwellsche Gleichung*
$\operatorname{rot}\boldsymbol{E}=-\partial\boldsymbol{B}/\partial t$.
Der Induktionsvorgang kann sowohl durch zeitliche Änderung der *magnetischen Flußdichte* $\int\boldsymbol{B}\cdot\mathrm{d}\boldsymbol{A}$ bei ruhender Leiterschleife als auch durch zeitliche Änderung des von der Leiterschleife erfaßten zeitlich konstanten Magnetfeldes

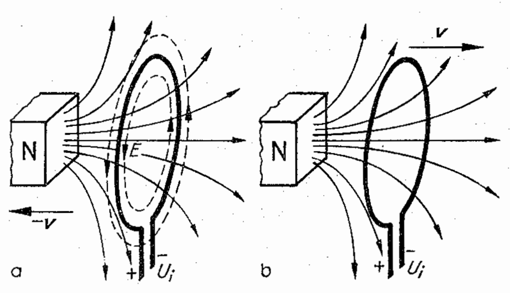


Abb. 2.4.4–3 Induktion (a) durch ein elektrisches Feld, das durch ein zeitlich veränderliches Magnetfeld aufgebaut wird, (b) durch Kraftwirkung auf eine im Magnetfeld bewegte Ladung

bei Relativbewegung der Leiterschleife (oder eines Teiles von ihr) zum Magnetfeld ausgelöst werden. Er äußert sich immer im Auftreten einer elektrischen Spannung. Die integrale Form der zweiten Maxwellschen Gleichung lautet

$$\oint E\cdot\mathrm{d}s=-\frac{\partial}{\partial t}\int B\cdot\mathrm{d}A\left(=-\int\frac{\partial \boldsymbol{B}}{\partial t}\mathrm{d}A+\oint(v\times B)\cdot\mathrm{d}s\right).$$

Das elektrische Feld wird durch ein zeitlich veränderliches Magnetfeld induziert, unabhängig davon, ob sich ein Leiter im Magnetfeld befindet oder nicht. Ist eine Leiterschleife vorhanden, so läßt sich die Induktion durch einen Spannungsstoß $\Phi=\int U_i\,\mathrm{d}t=\int\boldsymbol{B}\cdot\mathrm{d}\boldsymbol{A}$ nachweisen. Die Einheit von Φ ist Weber, 1 Wb = 1 Vs. Die magnetische Kraftflußdichte wird in Tesla gemessen, 1 T = 1 Vs/m².
In der Anordnung von Abb. 2.4.4–3 werden die Elektronen in Abb. 2.4.4–3a durch die magnetische Kraft $e^-(\boldsymbol{v}\times\boldsymbol{B})$, in Abb. 2.4.4–3b durch die elektrische Kraft $e^-\boldsymbol{E}$ bewegt. Der Unterschied zwischen beiden Experimenten läßt sich auch durch einen Übergang in ein anderes, in ein mit der Geschwindigkeit $\boldsymbol{v}$ bewegtes Bezugssystem verstehen. In der Relativitätstheorie (vgl. 3.) wird besprochen, wie sich elektrisches und magnetisches Feld ineinander transformieren. Sie sind also nicht als zwei wesensfremde Felder aufzufassen, sondern Teile des elektromagnetischen Feldes.
Die zur Induktion führende zeitliche Änderung von $\boldsymbol{B}$ kann bei zeitlich konstanter Feldstärke $\boldsymbol{H}$ durch eine zeitliche Änderung der relativen Permeabilität $\mu_r(t)$ des den Raum erfüllenden Stoffes und bei gleichbleibender magnetischer

Eigenschaft des Raumes mit der Feldstärke $\boldsymbol{H}(t)$ zeitlich geändert werden.
Im Wechselstromkreis wird $\boldsymbol{H}$ durch Änderung der Stromstärke zeitabhängig, es kommt in einer Spule zur Selbstinduktion, $U_i = -L\mathrm{d}I/\mathrm{d}t$. L ist die *Induktivität* der Spule, ihre Einheit ist Henry, 1 H = 1 Vs/A = 1 Wb/A. In einer als Toroid gewickelten Spule, d.h. einer Spule ohne magnetisches Streufeld, setzt sich die induzierte Spannung U_i aus den gleichgroßen Anteilen in den einzelnen n Windungen zusammen. Mit A als Spulenquerschnitt und l als Länge des Toroids ist $L = \mu_r\mu_0 n^2 A/l$. Durchsetzt der magnetische Fluß eine zweite Spule ganz oder teilweise, so sind beide induktiv gekoppelt. Die in der Spule 2 induzierte Spannung wird durch die Stromänderung in der Spule 1 bewirkt, $U_2 = -L_{12}(\mathrm{d}I_1/\mathrm{d}t)$. Dabei ist $L_{12} = L_{21}$ die *Gegeninduktivität* der beiden Spulen. Ihre Kopplung wird durch einen Kopplungsfaktor ausgedrückt, $L_{12} = k\sqrt{\mathrm{L_1L_2}}$ mit $k \leqq 1$. Für alle Fälle von Induktion gilt die *Lenzsche Regel*: Ein durch Induktion hervorgerufener Strom ist stets so gerichtet, daß er mit seinem Magnetfeld die Zustandsänderung zu hemmen sucht. Die Induktion in einem einfach zusammenhängenden Gebiet leitenden Materials führt zu *Wirbelströmen*.

2.4.4.2. Energiedichte des magnetischen Feldes. In einem Stromkreis aus Spannungsquelle U, Ohmschem Widerstand R und Induktivität L fließt nach dem Ohmschen Gesetz ein Strom $I = (U - L\mathrm{d}I/\mathrm{d}t)/R$. Die Lösung dieser Differentialgleichung lautet $I = U(1 - \mathrm{e}^{-Rt/L})/R$ mit $I = 0$ für $t = 0$. Beim Einschalten ist $t \ll L/R$, d.h. $I = Ut/L$. Das zeitliche Anwachsen des Stromes wird durch Spannung und Induktivität bestimmt und erst, wenn der Einfluß von $\mathrm{e}^{-Rt/L}$ mit wachsender Zeit immer geringer wird, kommt es zu einer Strombegrenzung durch den Widerstand, $I = U/R$. Wird der Stromkreis unterbrochen, bricht das Magnetfeld in der Induktivität zusammen und erzeugt durch Induktion nach der Lenzschen Regel eine der ursprünglichen Spannung gleichgerichtete Spannung, die einen Induktionsstrom $I = (U/R)\mathrm{e}^{-Rt/L}$ über einen am Schalter sich ausbildenden Lichtbogen abfließen läßt. Die gesamte im Magnetfeld der Induktivität L steckende Energie wird durch den Induktionsstrom abgeführt (vgl. 2.4.3.2.),

$$W = \int_0^\infty I_0^2 \mathrm{e}^{-2Rt/L} R\mathrm{d}t = LI_0^2/2.$$

Sie kann wegen $L = \mu n^2 A/l$ und $H = nI/l$ auch durch das Produkt aus $BH/2$ und Al ausgedrückt werden. Allgemein gilt für die Energiedichte des magnetischen Feldes $w = \boldsymbol{B}\cdot\boldsymbol{H}/2$.

2.4.4.3. Der Wechselstromkreis, Wechselspannungsquellen. Bei der Darstellung eines Wechselstromes $I(t) = I_0 \sin(\omega t + \varphi_I) = I_0 \sin\alpha$ bedeuten $I(t)$ den Momentanwert des Stromes, $\alpha = \omega t + \varphi_I$ den Phasenwinkel und φ_I den Nullphasenwinkel oder die Anfangsphase. Analog lautet die Spannung $U(t) = U_0 \sin(\omega t + \varphi_U)$. $\omega = 2\pi f = 2\pi/T$ ist die Kreisfrequenz mit f als Frequenz und T als Periodendauer. Ein sinusförmiger Wechselstrom wird einwellig genannt. Die wirklichen Wechselspannungen und -ströme sind zwar periodisch, weichen aber i. allg. von der Sinusform ab. Da sie nach FOURIER als Summe von sinusförmigen Spannungen bzw. Strömen darstellbar sind, werden sie mehrwellig genannt. Ihre Frequenzen sind ganzzahlige Vielfache einer Grundfrequenz oder der ersten Harmonischen. Kommt in der Fourier-Summe ein konstantes Glied vor, stellt es den Gleichspannungs- bzw. -stromanteil dar.
Der *Effektivwert* dient allgemein als Maß für die Stromstärke bzw. die Spannung eines beliebig periodischen Wechselstromes. Er entspricht dem Gleichstromwert, der in einem Ohmschen Widerstand die gleiche Wärmewirkung erzielen würde. Der über die positive Halbwelle erstreckte zeitliche Mittelwert der Momentanwerte ist der *elektrolytische Mittelwert*. Er entspricht dem Gleichstromwert bei der Elektrolyse mit gleichgerichtetem Wechselstrom. Bei sinusförmigem Verlauf ist der Effektivwert das $1/\sqrt{2}$- oder 0,707fache, der elektrolytische Mittelwert das $2/\pi$- oder 0,637fache des Scheitelwertes. Bei sinusförmigem Verlauf von Strom und Spannung ist die *elektrische Leistung des Wechselstromes* als *Wirkleistung* gleich dem zeitlichen Mittelwert der Momentanleistung während einer Periode darstellbar und gleich dem Produkt aus den Effektivwerten von Strom und Spannung und dem sogenannten Leistungsfaktor $\cos\varphi$: $\bar{N} = I_{\mathrm{eff}}U_{\mathrm{eff}}\cos\varphi$, wobei $\varphi = \varphi_I - \varphi_U$ die Phasenverschiebung zwischen Spannung und Strom ist. $I_{\mathrm{eff}}U_{\mathrm{eff}}\sin\varphi$ ist die *Blindleistung* und $I_{\mathrm{eff}}U_{\mathrm{eff}}$ die *Scheinleistung*.
Das bei Gleichstrom gültige Ohmsche Gesetz $U = IR$ gilt auch für die Augenblickswerte von Strom und Spannung im Stromkreis mit Kapazität bzw. Induktivität. Beim Ohmschen Widerstand sind Spannung und Strom in Phase $\varphi = \varphi_U - \varphi_I = 0$. Aus $U = Q/C$ und $I = \mathrm{d}Q/\mathrm{d}t$ folgt bei einer Kapazität $U_{\mathrm{eff}} = I_{\mathrm{eff}}/(\omega C)$ und $\varphi = \varphi_U - \varphi_I = -90°$. Der Strom eilt der Spannung um 90° oder $\pi/2$ voraus. Die Wirkleistung ist Null, die Blindleistung $-\omega CU_{\mathrm{eff}}^2$. Wegen $U + U_i = 0$ bzw. $U = L\mathrm{d}I/\mathrm{d}t$ gilt bei einer Induktivität $U_{\mathrm{eff}} = \omega LI_{\mathrm{eff}}$ und $\varphi = \varphi_U - \varphi_i = 90°$. Die Spannung eilt dem Strom voraus. Die Wirkleistung verschwindet wie bei der Kapazität, die Blindleistung ist $U_{\mathrm{eff}}^2/(\omega L)$. Sind die Spulenverluste durch den Spulendrahtwiderstand und gegebenenfalls durch Ummagnetisierung des Eisenkernes nicht vernachlässigbar, so kann die verlustbehaftete Spule im Ersatzschaltschema durch Induktivität L und Widerstand R in Reihe dargestellt werden.
Die Berechnung des Gesamtwiderstandes in einem Wechselstromkreis erfolgt zweckmäßig mit der komplexen Wechselstromrechnung. Der Ohmsche Widerstand R hat nur einen Realteil R, während der induktive und der kapazitive Widerstand nur einen Imaginärteil $\mathrm{j}\omega L$ bzw. $-\mathrm{j}/\omega C$ mit j als imaginärer Einheit $\mathrm{j} = \sqrt{-1}$ haben. Resultierender komplexer Widerstand bei einer Reihenschaltung ist $Z = Z_1 + Z_2$, bei Parallelschaltung errechnet er sich aus $Z = Z_1Z_2/(Z_1 + Z_2)$.
Zwei galvanisch getrennte Wechselstromkreise können durch einen *Transformator* induktiv gekoppelt werden. Er besteht im Prinzip aus Primär- und Sekundärwicklung in Form von zwei Spulen mit gemeinsamem Eisenkern. Beim idealen Transformator (Kopplungsfaktor $k = 1$, $L_{12} = \sqrt{L_1L_2}$, keine Verluste durch Streufelder) bestimmt das Verhältnis der Windungszahlen von Primär- zu Sekundärwicklung in einfacher Weise die Spannungs- und Stromübersetzung, $n_1/n_2 = -U_1/U_2 = -I_2/I_1$. Wegen der Sättigungserscheinung im Transformatorkern ist der Transformator ein nichtlinearer Widerstand im Wechselstromkreis. Die Wirbelstromverluste beim Ummagnetisieren versucht man durch Lamellierung des Eisenkernes bzw. bei hohen Frequenzen durch Verwendung von Eisenpulverkernen (vgl. 2.4.2.5., Ferrite) niedrig zu halten. Bei hohen Frequenzen sind unabhängig von der Art des Transformators die Wicklungskapazitäten zu berücksichtigen.

Wechselspannungsquellen dienen überwiegend zur Erzeugung einphasiger Wechselströme aus mechanischer Energie. Als Wechselstromdynamo bzw. Synchronmaschine bestehen sie aus einem rotierenden Polrad R (Rotor) mit

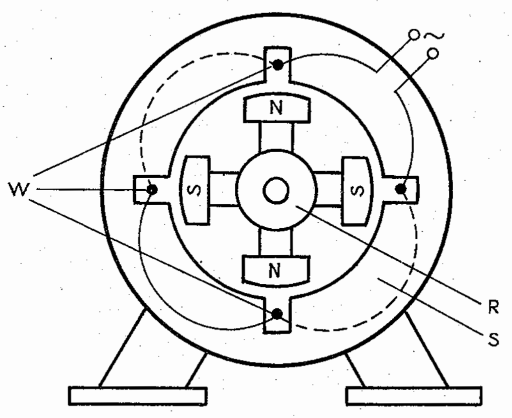


Abb. 2.4.4–4 Einphasengenerator (Wechselstromdynamo)

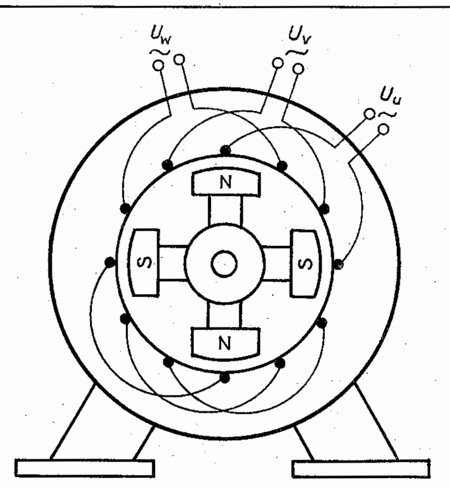


Abb. 2.4.4–5 Drehstromgenerator

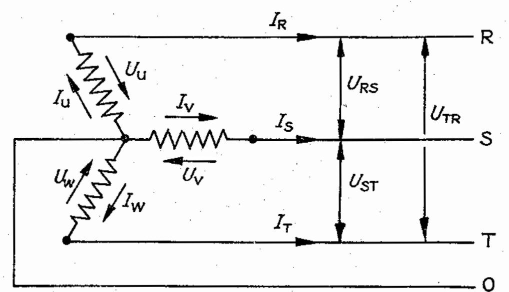


Abb. 2.4.4–6 Drehstromsystem in Sternschaltung

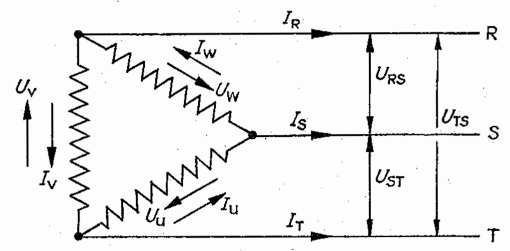


Abb. 2.4.4–7 Drehstromsystem in Dreieckschaltung

abwechselnd aufeinanderfolgenden Nord- und Südpolen und einem äußeren, feststehenden Anker S (Stator) (Abb. 2.4.4–4). Zur Erzeugung des magnetischen Kraftflusses im Polrad dient eine in der Abbildung nicht gezeichnete Erregerwicklung, die in Form von Spulen die Pole umgibt. Der durch diese Wicklungen fließende Gleichstrom (Erregerstrom) wird dem Rotor über Schleifringe von außen zugeführt. Der Stator enthält in Nuten die Wechselstromwicklung W, deren Verbindung auf der Rückseite gestrichelt dargestellt ist. Die aus den Nordpolen austretenden magnetischen Kraftlinien gelangen durch den Luftspalt zwischen Rotor und Stator in das Statoreisen, verlaufen dort nach beiden Seiten, um über den Luftspalt in die beiden dem Nordpol benachbarten Südpole des Rotors zurückzutreten und sich im Eisen des Rotors zu schließen. Das mit dem Polrad bewegte Magnetfeld induziert in der Statorwicklung die für die Erzeugung des Wechselstromes erforderliche Wechselspannung. Mit der Zahl der Polpaare p des Rotors und seiner Drehzahl n je Minute ist die Frequenz des erzeugten Wechselstromes $f = (pn/60)\mathrm{s}^{-1}$.

Bringt man im Stator zwei weitere derartige Wicklungen an (Abb. 2.4.4–5), die aber gegen die erste und gegeneinander um 2/3 Polteilung versetzt sind, erhält man in diesen Wicklungen drei Wechselspannungen U_u, U_v und U_w, die in der Phase gegeneinander um je 1/3 Periode verschoben sind. Den in den drei Leitersystemen fließenden Strom bezeichnet man als *Drehstrom.* In Sternschaltung haben die drei Leitersysteme einen gemeinsamen Nulleiter und drei Außenleiter, in Dreieckschaltung sind die Wicklungsenden von je zwei aufeinander folgenden Statorwicklungen zu je einem Leiter zusammengefaßt. Für die Effektivwerte in einem Drehstromleitersystem gilt bei Sternschaltung $U_{RS} = U_{ST} = U_{TR} = \sqrt{3}\, U_u = \sqrt{3}\, U_v = \sqrt{3}\, U_w$, Netzspannung $= \sqrt{3}$ Phasenspannung und Netzstrom = Phasenstrom (Abb. 2.4.4–6); bei einer Dreieckschaltung gilt $U_{RS} = U_{ST} = U_{TR} = U_u = U_v = U_w$, Netzspannung = Phasenspannung und der Netzstrom $= \sqrt{3}$ Phasenstrom (Abb. 2.4.4–7). Leitet man dem Stator einer Drehstrommaschine einen an anderer Stelle erzeugten Drehstrom zu, werden durch die drei Wicklungen im Stator Magnetfelder zeitlich und räumlich aufeinanderfolgend erzeugt, die gemeinsam ein *Drehfeld* bilden. In diesem läuft ein Polrad synchron, ein Kurzschlußanker asynchron mit und gibt mechanische Leistung durch ein Drehmoment ab, so daß das Produkt aus Drehmoment und Winkelgeschwindigkeit zuzüglich der Verlustleistung gleich der dem Netz entnommenen elektrischen Leistung ist.

2.4.4.4. Elektromagnetische Schwingungen. Jede Reihen- und jede Parallelschaltung von Induktivität und Kapazität stellt einen Schwingkreis mit der Eigenfrequenz $f = (1/2\pi)\sqrt{1/(LC)}$ dar. Verluste durch Joulesche Wärme in den Zuleitungen und in den Drahtwindungen von Spulen, beim Ummagnetisieren des Spulenkernes bzw. Umpolarisieren des Dielektrikums in den Kondensatoren werden summarisch durch einen Ohmschen Widerstand dargestellt. Solange $R < 2\sqrt{L/C}$ gilt, kommt es zu gedämpften Schwingungen, anderenfalls zu einem aperiodischen Abfall der am Anfang aufgeprägten Schwingungsamplitude. $R = 2\sqrt{L/C}$ nennt man den *Grenzfall.*

Ungedämpfte elektromagnetische Schwingungen lassen sich nur erzeugen, wenn die Verluste durch Zufuhr elektrischer Energie im Rhythmus der Schwingungen ausgeglichen werden. Ein klassisches Beispiel hierfür ist die Meissnersche Rückkopplungsschaltung (Abb. 2.4.4-8). Das Auskoppeln der elektromagnetischen Schwingungen aus einem Schwingkreis kann induktiv, konduktiv, kapazitiv und galvanisch erfolgen.

Durch Verkleinern der Induktivität zu einer „Spule“ mit nur einer Windung und durch Verkleinern der Kapazität zu einem „Kondensator“ aus zwei einander gegenüberliegenden Drahtenden nimmt der Schwingkreis die Form eines offenen Drahtringes an. Elektrisches und magnetisches Feld sind nicht mehr wie beim Schwingkreis getrennt in Kondensator und Spule lokalisiert. Auch wenn der Drahtring aufgebogen wird, hat er noch eine Kapazität und eine Induktivität. Er wird als gerades Leiterstück zum linearen oder Hertzschen Oszillator, dessen elektromagnetische Schwingungen an den Leiterenden Stromknoten und in der Mitte des Leiters einen Spannungsknoten haben. Sie können als eine stehende Welle aufgefaßt werden. Die Resonanzfrequenz des linearen Oszillators ist $f = c/2l$, wobei die Phasengeschwindigkeit c der Welle der Lichtgeschwindigkeit gleichgesetzt werden kann. l ist die Länge des Oszillators. Das Resonanzfrequenzmaximum ist wegen der Strahlungsdämpfung sehr breit. Der Oszillator gleicht einem Dipol, dessen elektrisches und magnetisches Feld sich periodisch ändern. Vom Dipol löst sich ein Wechselfeld ab. Bereits im Abstand von wenigen Wellenlängen hat das sich in den Raum ausbreitende elektromagnetische Wechselfeld die Eigenschaften einer Transversalwelle. Die elektrische und der magnetische Feldvektor schwingen phasengleich, stehen senkrecht zur Ausbreitungsrichtung und senkrecht aufeinander.

2.4.4.5. Elektromagnetische Wellen. Das Spektrum der elektromagnetischen Wellen reicht von den γ-Strahlen mit einem Wellenlängenbereich von 10^{-11} bis 10^{-16} m bis zu den mit Röhrensendern erzeugten sogenannten Längstwellen mit Wellenlängen von 5 bis 20 km.

Die Ausbreitung der elektromagnetischen Wellen ist nicht an ein Übertragungsmedium gebunden, im stofferfüllten Raum aber je nach Wellenlängenbereich sehr unterschiedlich. Die durch eine bestimmte Feldstärke am Empfangsort definierbare Reichweite eines Rundfunksenders im Mittel- und Langwellenbereich bei gleichbleibender Sendeleistung wächst mit zunehmender Wellenlänge und mit größerer elektrischer Leitfähigkeit des Erdbodens. Im UKW-Bereich nimmt die Reichweite außerdem durch Erhöhen der Sende- und der Empfangsantenne zu. Die Zunahme durch erhöhte Aufstellung ist aber um so größer, je geringer die Bodenleitfähigkeit ist. Bei Mittel- und Kurzwellen nimmt die Feldstärke der Bodenwelle wegen der mit der Frequenz zunehmenden Dämpfung bei wachsender Entfernung um so schneller ab, je kürzer die Wellenlänge ist. Der Bodenwelle überlagern sich dann in zunehmendem Maße die Raumwellen, die an leitenden Schichten der Ionosphäre reflektiert werden. Bei noch kürzeren Wellenlängen im dm-Bereich kommt es nicht mehr zur Reflexion an der Ionosphäre.

Von einer Ausbreitung elektrischer Wellen auf Drähten kann man sprechen, wenn an einen Empfangsdipol eine *Paralleldrahtleitung* angeschlossen ist. Strom und Spannung genügen einer Wellengleichung (Abb. 2.4.4-9): Bei 1 tritt in den Leiterabschnitt dx die Ladung $I\mathrm{d}t$ ein, bei 2 die Ladung $\left(I + \frac{\partial I}{\partial x}\mathrm{d}x\right)\mathrm{d}t$ aus. Mit der im Leiterabschnitt dx verbleibenden Ladung $-\frac{\partial I}{\partial x}\mathrm{d}x\mathrm{d}t$ stellt sich zwischen den Leitern aufgrund der Kapazität je Längeneinheit C' eine Querspannung $\mathrm{d}U = -\frac{\partial I}{\partial x}\mathrm{d}x\mathrm{d}t/C'\mathrm{d}x$ ein, woraus $\frac{\partial U}{\partial t} = -\frac{1}{C'}\frac{\partial I}{\partial x}$ folgt. Entsprechend ist bei einer Selbstinduktivität der Doppelleitung je Längeneinheit L' die durch die Stromänderung induzierte Gegenspannung $\mathrm{d}U = -L'\mathrm{d}x\frac{\partial I}{\partial t}$, woraus $\frac{\partial U}{\partial x} = -L'\frac{\partial I}{\partial t}$ folgt. Nach weiterer partieller Ableitung nach x bzw. t und Gleichsetzen der rechten Seiten folgt die Wellengleichung für I,

$$\frac{\partial^2 I}{\partial x^2} = L'C'\frac{\partial^2 I}{\partial t^2}.$$

In entsprechender Weise läßt sich die Wellengleichung für U ableiten. $\sqrt{L'/C'}$ ist der Wellenwiderstand Z der Doppelleitung. Bei offenem und bei kurzgeschlossenem Ende der Doppelleitung kommt es zu stehenden Wellen durch Reflexion mit um $\lambda/4$ gegeneinander versetzten Strom- und Spannungsknoten bzw. -bäuchen. Haben die Leitungen einen nicht vernachlässigbaren Ohmschen Widerstand je Längeneinheit R' und durch ungenügende Isolation zwischen den beiden Leitern eine Ableitung je Längeneinheit G', sind die sich auf der Paralleldrahtleitung ausbreitenden Wellen gedämpft, und es kommt zur Dispersion. Für den Wellenwiderstand gilt dann $Z = \sqrt{(R' + \mathrm{j}\omega L')/(G' + \mathrm{j}\omega C')}$ mit j als imaginärer Einheit und ω als Kreisfrequenz. Wird eine Paralleldrahtleitung mit ihrem Wellenwiderstand abgeschlossen, verhält sie sich in

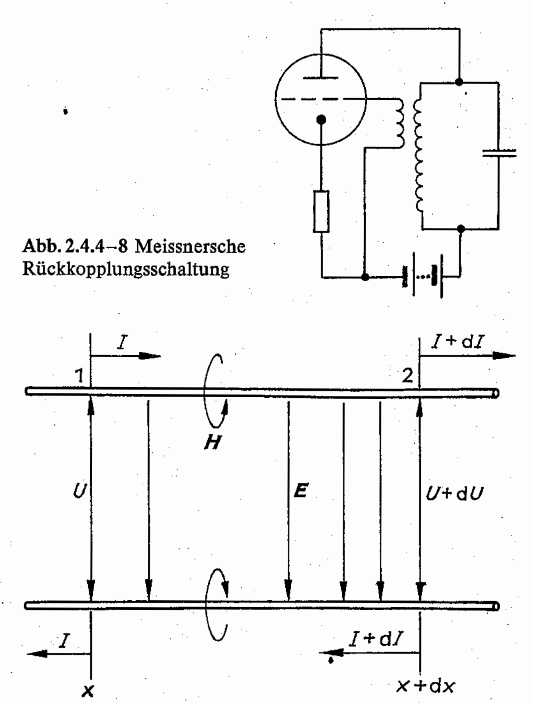


Abb. 2.4.4–8 Meissnersche Rückkopplungsschaltung

Abb. 2.4.4–9 Zur Ableitung der Wellengleichung für elektromagnetische Schwingungen auf einer Paralleldrahtleitung

bezug auf die Ausbreitung von Wellen wie eine unendlich lange Leitung, es kommt nicht zur Reflexion. Eine spezielle Form der Paralleldrahtleitung ist das *Koaxialkabel.* In *Hohlleitern* erregte elektrische oder magnetische Felder breiten sich in Wellenform erst oberhalb einer Grenzfrequenz aus.

Die elektromagnetischen Wellen in Hohlleitern sind keine reinen Transversalwellen! Ihre Wellenlänge λ_H ist immer größer, als die von elektromagnetischen Wellen gleicher Frequenz im freien Raum, $\lambda_H = \lambda/\sqrt{1-(\lambda/\lambda_k)^2}$. Um denselben Faktor ist auch ihre Phasengeschwindigkeit größer als die Lichtgeschwindigkeit im Vakuum. λ_k ist dabei eine Grenzwellenlänge, die von der Symmetrie und der Form des Hohlleiterquerschnittes mit seinem Durchmesser D abhängt und immer kleiner als $2D$ ist.

Bei Paralleldrahtleitern und Koaxialleitern ist die Phasengeschwindigkeit u der elektromagnetischen Wellen bei Einbettung in bzw. Isolierung durch ein homogenes Dielektrikum kleiner als die Lichtgeschwindigkeit im Vakuum, $u = c/\sqrt{\varepsilon_r/\mu_r}$.

In gewissem Sinne ist ein beidseitig abgeschlossener Hohlleiter als Hohlraumresonator die Umkehrung des linearen Oszillators. Die in ihm erzeugbaren stehenden Wellen werden nicht abgestrahlt, es kommt nicht zu einer Strahlungsdämpfung wie beim linearen Oszillator; es treten nur scharfe Resonanzmaxima auf.

2.5. Optik

Unter dem Begriff Optik wird alles verstanden, was mit den Gesetzmäßigkeiten der Abbildung, mit der Natur des Lichtes, das diese Abbildung u. a. visuelle Eindrücke erst ermöglicht, und mit dem Sehvorgang als solchem, d. h. mit der Wahrnehmung unserer Umwelt durch den Gesichtssinn, zusammenhängt. Der Strahlen-, Wellen- und Quantenoptik, die weitgehend auf eine von der Wahrnehmung, dem Sehen unabhängige Grundlage gestellt sind, wird eine, in Anlehnung an den Titel eines Hauptwerkes von HELMHOLTZ, physiologisch genannte Optik an die Seite gestellt. In ihr gelten die an den menschlichen Sinnen und der Psyche begründeten Gesetze. Alle Lebensbereiche werden von ihr berührt. Der ursprünglich der Optik zukommende sichtbare Spektralbereich der elektromagnetischen Wellen ist heute erweitert auf die angrenzenden Bereiche des ultravioletten und infraroten Lichtes.

2.5.1. Das Licht

Licht ist von Atomen und Molekülen als Folge von Zustandsänderungen abgestrahlte elektromagnetische Energie. Parallel engbegrenzte Lichtbündel oder -strahlen werden für den Gebrauch in der geometrischen bzw. Strahlenoptik zu geometrischen Linien vereinfacht, die den Weg des Lichtes darstellen sollen. In der physikalischen oder Wellenoptik wird das Licht als elektromagnetische Wellen aufgefaßt, den Rundfunkwellen wesensgleich. Unterschiede bestehen in der Wellenlänge, der Wechselwirkung mit der Materie und vor allem in der Erzeugung. Sich in einer Richtung ausbreitendes Licht enthält stets eine sehr große Anzahl einzelner Wellenzüge mit derselben Ausbreitungsrichtung. Es wird monochromatisch genannt, wenn es nur Wellenzüge einer oder nahezu einer Wellenlänge enthält. Licht ist eine transversale Welle (vgl. 2.3.1.3.), der Vektor des elektrischen Feldes E steht senkrecht auf der Ausbreitungsrichtung. Licht ist unpolarisiert, wenn die Orientierung der E-Vektoren seiner einzelnen Wellenzüge auf alle Richtungen senkrecht zur Ausbreitungsrichtung gleichverteilt ist. Es ist linear polarisiert, wenn die Orientierung der E-Vektoren seiner einzelnen Wellenzüge auf eine Richtung senkrecht zur Ausbreitungsrichtung beschränkt ist. Die von der Ausbreitungsrichtung und den E-Vektoren aufgespannte Ebene wird *Schwingungsebene* genannt, im Gegensatz zu dem nur historisch zu verstehenden Begriff *Polarisationsebene.* Diese steht auf der Schwingungsebene senkrecht. Licht kann auch zirkular bzw. elliptisch polarisiert sein (vgl. 2.3.1.3.). In der Quantenoptik wird (vgl. 4.11.2.) das Licht als ein Photonenstrom aufgefaßt, der der Teilchennatur des Lichtes Rechnung trägt.

2.5.1.1. Lichtsinn und Fotometrie. Unter Lichtsinn versteht man den Sinn, der durch die Reaktion von Sinneszellen auf Lichteinwirkung erregt wird. Im menschlichen Auge sind die für den Lichtsinn spezialisierten primären Sinneszellen die Stäbchen und Zapfen in der Netzhaut. Auf ihr wird über den dioptrischen Apparat des Auges (vgl. 2.5.3.5.) die Umwelt abgebildet. Der Bildpunkt eines beim Sehen fixierten Objektpunktes liegt in der Netzhautgrube. Sie enthält nur die für Tag- bzw. Farbensehen spezialisierten Sinneszellen, die Zapfen, und ist sozusagen nachtblind. Bei fortgeschrittener Dunkeladaptation scheint ein mit der Netzhautperipherie wahrgenommenes lichtschwaches Objekt, z. B. ein Stern, zu verschwinden, wenn man den Blick bewußt darauf richtet. Eine völlige strukturelle und funktionelle Trennung der zu den Stäbchen und den Zapfen gehörenden nachgeschalteten neuronalen Systeme in der Netzhaut gibt es nicht. Damit läßt sich der gleitende Übergang vom Dämmerungs- zum Tagsehen und umgekehrt verstehen. Bei Helladaptation werden in der Netzhaut durch neuronale Signalverarbeitung Kontrastmechanismen eingeleitet, der Simultankontrast als Überhöhung des Hell-Dunkel an einer Hell-Dunkel-Grenze im Bild auf der Netzhaut sowie der Sukzessivkontrast als kurzdauerndes Nachbild in der ersten Sekunde. Der Gegensatz dazu sind die langandauernden Nachbilder, die treffend auch *Blendungsbilder* genannt werden. Beide Kontrastmechanismen verlieren sich ganz mit vollständiger Dunkeladaptation. In der Augenlinse wird Licht mit einer Wellenlänge von über 830 nm und weniger als 360 nm vollständig absorbiert. Bei Linsenlosigkeit infolge Staroperation ist Lesefähigkeit und damit Gesichtssinn bei Licht bis zu Wellenlängen um 300 nm nachgewiesen worden. Die Hellempfindlichkeit des Auges ist in den einzelnen Spektralbereichen des sichtbaren Lich-

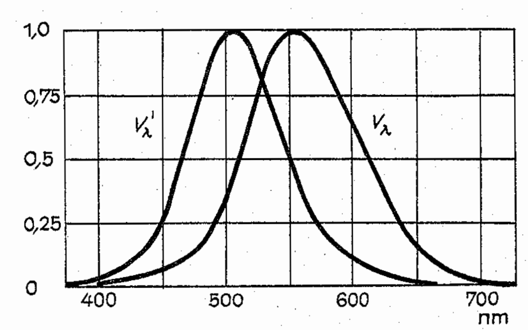


Abb. 2.5.1-1 Spektraler Hellempfindungsgrad des Auges

tes verschieden. Sie ist für jede Wellenlänge durch den Quotienten aus dem fotometrisch gemessenen Lichtstrom der entsprechenden monochromatischen Strahlung und seiner physikalisch gemessenen Leistung definiert. Bezogen auf den gleich 1,0 gesetzten Maximalwert ist sie der spektrale Hellempfindungsgrad des Auges V_λ (Abb.2.5.1–1). Gegen die Wellenlänge aufgetragen erhält man eine Tages- und eine Nachtwertkurve, je nachdem, ob das Auge hell- oder dunkeladaptiert ist. Die entsprechenden V_λ- bzw. V'_λ-Werte für beide Kurven mit ihren Maximalwerten 1,0 bei 555 nm bzw. bei 507 nm sind international festgelegt und auf einen Normalbeobachter bezogen.

Bei der *visuellen Fotometrie* wird die Lichtstärke bzw. die Leuchtdichte einer Lichtquelle durch Vergleich mit einem Normal bestimmt. Das Auge dient dabei als Nullinstrument zur Kontrolle der Einstellung auf gleiche von beiden Lichtquellen in einem Fotometer erzeugte Beleuchtungsstärke bzw. Leuchtdichte. Bei der *physikalischen Fotometrie* muß der für den Vergleich benutzte Empfänger dem spektralen Hellempfindungsgrad des Auges V_λ angepaßt sein. Sind die Abmessungen der zu vergleichenden Lichtquellen klein gegen die Entfernungen von dem Ort, an dem die von ihnen erzeugte Beleuchtungsstärke bzw. Leuchtdichte als gleich empfunden oder gemessen wird, dann ist das Verhältnis ihrer Lichtstärken I_i gleich dem Verhältnis der Quadrate ihrer Entfernungen r_i vom Vergleichsort, $I_1/I_2 = r_1^2/r_2^2$. Bei Lichtquellen mit Scheinwerferwirkung gilt dieses Abstandsgesetz erst für eine Entfernung r_i, aus der die jeweilige Lichtquelle gleichmäßig leuchtend erscheint. Alle fotometrischen Maßgrößen beziehen sich auf die Basiseinheit candela: 1 Candela (cd) ist die Lichtstärke senkrecht zu 1/600 000 m² der Oberfläche eines schwarzen Strahlers bei der Temperatur des beim Druck von 101325 Pa erstarrenden Platins. Eine punktförmige Lichtquelle der *Lichtstärke* 1 cd strahlt in den Raumwinkel 1 sr den *Lichtstrom* 1 Lumen (lm) ab, 1 lm = 1 cd · sr. Auf einer Fläche von 1 m² erzeugt ein gleichmäßig verteilter Lichtstrom von 1 lm die *Beleuchtungsstärke* 1 Lux (lx), 1 lx = 1 cd · sr · m⁻². Damit gleichbedeutend ist die Aussage, daß eine punktförmige Lichtquelle der Lichtstärke 1 cd in 1 m Entfernung die Beleuchtungsstärke 1 lx erzeugt. Die Oberfläche eines schwarzen Strahlers hat bei der Temperatur des erstarrenden Platins definitionsgemäß die *Leuchtdichte* 60 Stilb (sb), 1 sb = 1 cd/cm². Eine alles Licht diffus reflektierende mattweiße Fläche hat bei einer Beleuchtungsstärke von 1 lx die Leuchtdichte 1 Apostilb (asb). Bei der Beleuchtungsstärke von 1 Phot (ph) ist die Leuchtdichte einer solchen Fläche 1 Lambert (la), $1\,\text{asb} = 1/(\pi \cdot 10^4)\,\text{cd/cm}^2$, $1\,\text{ph} = 1\,\text{lm/cm}^2$, $1\,\text{la} = 1/\pi\,\text{cd/cm}^2$. Die Einheit Phot wird sowohl bei der Angabe von sehr hohen Beleuchtungsstärken, wie sie bei der Abbildung von Lichtquellen auftreten, als auch zur Angabe der spezifischen Lichtausstrahlung benutzt. Zur Umrechnung physikalischer Strahlungsmaße in fotometrische bzw. lichttechnische dient das fotometrische *Strahlungsäquivalent* $K_\lambda = K_{max} \cdot V_\lambda$, wobei $K_{max} = (680 \pm 5)\,\text{lm/W}$ und $1/K_{max} = (1{,}47 \pm 0{,}01) \cdot 10^{-3}\,\text{W/lm}$ das *Lichtäquivalent M* ist. Bei Leuchtdichten von weniger als $5 \cdot 10^{-2}$ asb liegt reines Stäbchensehen vor, für die Hellempfindlichkeit des Auges gilt die Nachtwertkurve V'_λ. Um bei der fotometrischen Aussage den Fall des Stäbchensehens gegen den Normalfall des helladaptierten Auges abgrenzen zu können, sind für die Beleuchtungsstärke die Einheit 1 Nox (nx) und für die Leuchtdichte die Einheit 1 Skot (sk) eingeführt worden, $1\,\text{nx} = 10^{-3}\,\text{lx}$ und $1\,\text{sk} = 10^{-7}/\pi\,\text{sb}$. Während bei helladaptiertem Auge Leuchtdichten von $5 \cdot 10^{-2}$ asb bis zur Blendungsschwelle bei 10^6 asb wahrgenommen werden, ist bei dunkeladaptiertem Auge eine Lichtwahrnehmung von der Absolutschwelle bei $5 \cdot 10^{-6}$ asb bis zu 10 asb möglich. Es ist jedoch nicht sinnvoll, bei Fotometrie mit normallichttüchtigem Auge Leuchtdichten unter 10 sk anzugeben, da dann immer mit einem Anteil Zapfensehen gerechnet werden muß.

2.5.1.2. Farbensehen. Bei geschlossenem Auge und in einem total verdunkelten Raum entsteht die subjektive Wahrnehmung des sogenannten Eigengrau. Wird mit der Leuchtdichte die Absolutschwelle für Lichtwahrnehmung überschritten, setzt das Dämmerungssehen ein. Ein Glühlichtspektrum erscheint farblos. Das langwellige Ende des Spektrums ist verkürzt und reicht nur bis 640 nm. Das daran anschließende Rot, für das die Stäbchenrezeptoren blind sind, wird „dämmerwertfreies Rot" genannt. Die Extinktionskurve des in den Stäbchenrezeptoren wirksamen Sehpurpurs stimmt mit der Nachtwertkurve der Hellempfindlichkeit des Auges überein. Läßt man die Leuchtdichte des Spektrums anwachsen, bis die Helladaptation einsetzt, erkennt man zunächst nur drei Farben: Rot (zwischen 760 und 570 nm), Grün (zwischen 570 und 480 nm) und Blau (zwischen 480 und 380 nm). Während das zunächst farblose Grün und Blau über zunehmende Helligkeit in ihre Farbwirkung übergehen, tritt das dämmerwertfreie Rot farbig über die Wahrnehmungsschwelle. Mit zunehmender Leuchtdichte nimmt die Zahl der unterscheidbaren Farbtöne zu. Wird einem kontinuierlichen Spektrum gleichmäßig über seine ganze Länge unbuntes Licht hinzugefügt, treten die Farben Gelb, Blaugrün und Violett auffällig hervor. Als besonders rein werden die Spektralfarben der folgenden Wellenlängen empfunden: Rot 670, Orange 620, Gelb 600, Grün 520, Blau 470 und Violett 420 nm. Bei helladaptiertem Auge kann ein normalfarbtüchtiges Auge maximal 180 Spektralfarben unterscheiden. Mit den Purpurfarben erhöht sich diese Zahl auf 200. Oberhalb einer Leuchtdichte von 10^5 asb nimmt das Unterscheidungsvermögen für die Spektralfarben wieder ab. Bei Blendung können nur noch Blauviolett und Gelb mit hohem Weißanteil wahrgenommen werden. Ähnlich der Hell-Dunkel-Adaptation gibt es auch eine Farbadaptation, farbige Verstimmung oder Umstimmung genannt. Der Vorgang der Umstimmung wird etwa 1,5 s nach Wahrnehmung des Farbreizes (vgl. 2.5.1.4.) wirksam. Solange befindet sich das Auge in Neutralstimmung. Nach einer Einwirkungszeit von 2 min ändert sich der umgestimmte Zustand des Auges nicht mehr wesentlich. Im Spektrum gibt es drei Wellenlängen, bei denen keine farbige Umstimmung stattfindet. Sie liegen bei 560, bei 500 bis 490 und bei 460 nm. Bei Farbreizen zwischen dem kurzwelligen Ende des Spektrums und der mittleren Wellenlänge wird der Farbton in Richtung auf 460 nm, bei Farbreizen aus der anderen Hälfte des Spektrums in Richtung auf 560 nm durch Umstimmung verschoben. Beim Tragen von farbigen Brillen und in künstlich beleuchteten Räumen findet ebenfalls eine farbige Umstimmung statt. Die für Farbmessungen und Farbvergleiche notwendige Neutralstimmung des Auges wird durch Helladaptation an eine helle weiße Fläche oder an das Tageslicht ohne Fixierung farbiger Objekte bewirkt. Die farbige Umstimmung ist die Ursache für farbige Nachbilder im Sinne eines *Sukzessivkontrastes.* Ein auf eine Netzhautstelle einwirkender Farbreiz ruft einen farbigen *Simultankontrast* hervor. So erscheint ein unbunter Teil des Gesichtsfeldes in einer bunten Umgebung in der Kontrastfarbe. Die ursprüngliche, bei Neutralstimmung gesehene Farbe wird auch als induzie-

rende Farbe bezeichnet, die Kontrastfarbe als induzierte, bei GOETHE als geforderte: „Gelb fordert Rotblau, Blau fordert Rotgelb, Purpur fordert Grün und umgekehrt". Wenn man die Spektralfarbe Gelb, ohne ihre absolute Leuchtdichte zu ändern, relativ zu ihrer Umgebung verdunkelt, verliert sie bald den Charakter des Gelben und wird Olivgrün. Ein lichtes Orange wird bei dem gleichen Vorgang zu Braun, Weiß zu Grau und Rot zu Dunkelrot. Durch die Zugabe von Weiß wird Rot zu Rosa aufgehellt, Blau zu Hellblau. Beide Arten von Abwandlung einer Farbe werden als *Verhüllung* bezeichnet, obwohl es sich um zwei qualitativ unterschiedliche Vorgänge handelt. Bei dem einen wird durch die Zugabe von Weiß die spektrale Zusammensetzung des Farbreizes geändert, bei dem anderen beruht die Änderung des Farbeindruckes auf der Adaptation an das Umfeld bei gleichzeitigem Simultankontrast. Trotzdem wird die zusammenfassende Darstellung aller Verhüllungen eines Farbtones in einem *Verhüllungsdreieck* (vgl. Taf. 34) mit den Eckpunkten Weiß, Schwarz und unverhülltem Farbton als eine sehr naheliegende Systematik der Farben empfunden. Erscheinen im Gesichtsfeld zwei Farbreize gleichzeitig oder so schnell abwechselnd, daß die aufeinander folgenden Farbempfindungen miteinander verschmelzen, kommt es zu einer additiven Farbmischung. Es wird nur noch eine Farbe gesehen. Die additive Mischung aller Spektralfarben führt zum Farbeindruck Weiß. Als unbunt gilt ein Farbreiz, den eine Strahlung mit energiegleichem Spektrum hervorruft. Ihr entspricht eine *Farbtemperatur* (vgl. 2.5.1.5.) von 5270 K. Zum Farbeindruck Purpur kommt es bei additiver Mischung der Farben von den beiden Enden des Spektrums. Diese werden mit kontinuierlich sich änderndem Mischungsverhältnis durch eine kontinuierliche Folge von Purpurfarbtönen ineinander überführt. Zusammen mit den Spektralfarben lassen sie sich in einem Farbenkreis (Taf. 34) anordnen, in dem jeder Farbton seinen beiden Nachbarn ähnlicher ist als jedem anderen. Farben größter Unähnlichkeit liegen sich im Farbenkreis als Gegenfarben gegenüber. GOETHE nannte sie „polare Farben".

2.5.1.3. Körperfarben. Ein nichtselbstleuchtender Körper erhält seine Farbe durch wellenlängenabhängige Transmission bzw. Remission des auf ihn auftreffenden Lichtes. Bei einem klar transparenten Stoff entsteht die Farbe entsprechend dem Absorptionsspektrum der in ihm enthaltenen Farbzentren bzw. der in ihm gelösten Farbstoffmoleküle. Für jede Wellenlänge gilt dabei das *Lambert-Beersche Gesetz*, wonach der Absorptionskoeffizient der Konzentration c proportional ist (vgl. 2.5.2.2.). Die Farbenmischungen sind subtraktiv, da sich die Absorptionskoeffizienten addieren. Handelt es sich um die Einfärbung durch Farbstoffpigmente oder Kolloide, kommt zu der absorbierenden Wirkung eine streuende hinzu, die bei zunehmender Teilchenzahldichte in eine Vielfachstreuung übergeht. Die Farben opaker Stoffe beruhen in gleicher Weise auf selektiver Absorption und Remission des auf sie auftreffenden Lichtes. Von allen Körperfarben nimmt das Weiß eine besondere Stellung ein, insbesondere das vollkommen matte Weiß. In der Farbmessung dient eine frische MgO-Fläche als Normalweiß. Ihr *Remissionsgrad* ist Eins. Er ist gleich dem Verhältnis der spektralen Leuchtdichte einer Probe zu der einer unter den gleichen Beleuchtungsbedingungen stehenden vollkommen mattweißen Fläche. Das vollkommene Schwarz hat den Remissionsgrad Null. Es läßt sich durch die Öffnung zu einem lichtlosen Hohlraum darstellen. Körperfarben werden durch Vergleich mit Farbvorlagen bestimmt, wobei gleiche Beleuchtung mit einer *Normallichtart* vorausgesetzt ist. Tageslicht als Normallichtart D 65 wird mit Xenon-Hochdrucklampen realisiert bei einer *Farbtemperatur* 6500 K. Als Normallichtart A ist ein Planckscher Strahler bei 2856 K definiert, der sich mit Glühlampen darstellen läßt. Die *Hefner-Kerze* (HK) hat dagegen nur eine *Farbtemperatur* von 1930 K und wird bereits als bunt empfunden.

2.5.1.4. Farbenmetrik. Die Farbenmetrik ist die Lehre von den Farben als meßbare Größen. Ihr ist die Aufgabe gestellt, die Farbenwelt durch Maß und Zahl zu ordnen und für Farbunterschiede der Farbempfindung angepaßte Maße bereitzustellen. In diesem Sinne unterscheidet man eine niedere von einer höheren Farbenmetrik. Die niedere Farbenmetrik lehrt die Kennzeichnung der Farben zum Zweck ihrer Identifizierung, während die höhere Farbenmetrik im wesentlichen Formeln zur Farbunterschiedsbewertung und Maßgrößen für Farbabweichungen von dem Original oder der Vorlage bei der Wiedergabe durch Foto, Fernsehen, Druck usw. bereitzustellen hat. Die eine Farbempfindung auslösende Strahlung wird *Farbreiz*, der durch den Farbreiz hervorgerufene Farbeindruck die Farbe oder auch die *Farbvalenz* genannt. Zwei gleichfarbig erscheinende Farbreize sind isochrom, bei gleicher spektraler Zusammensetzung isomer oder unbedingt gleich, bei verschiedener spektraler Zusammensetzung metamer oder bedingt gleich. Auf das Ergebnis einer additiven Farbmischung hat nur das Aussehen der zu mischenden Farbreize und nicht deren spektrale Zusammensetzung Einfluß. Zur Kennzeichnung eines Farbreizes sind drei voneinander unabhängige Bestimmungsgrößen notwendig und hinreichend. Alle Farbmischungsreihen sind stetig. Bei additiver Farbmischung addieren sich die Leuchtdichten algebraisch. Zwischen vier verschiedenen Farbreizen läßt sich stets eine lineare Beziehung in Form einer Farbgleichung aufstellen. Die Möglichkeit einer Farbmessung ist dadurch gegeben, daß sich jeder beliebige Farbreiz zu drei Spektralreizen in eine lineare Beziehung setzen läßt. Einem Vorschlag der CIE (Commission Internationale d'Éclairage) folgend werden die *Spektralreize* mit den Wellenlängen 435,8, 546,1 und 700,0 nm als *reelle Eichreize* benutzt. In einer *Farbtafel* mit einem gleichseitigen *Farbendreieck*, in dem die Eckpunkte durch diese Eichreize dargestellt werden, liegen die Farborte entsprechend den Eichreizanteilen des Farbreizes gemessen in Dreieckskoordinaten. Bei linearen Dreieckskoordinaten liegt der *Weißpunkt W* nicht notwendig in der Mitte des Dreiecks. Die Spektralfarben bilden mit der Purpurgeraden als Verbindungslinie zwischen den Farborten der Enden des Spektrums den in sich geschlossenen Spektralfarbenzug. Bei den Spektralfarben und den Farben mit Farborten zwischen Spektralfarbenzug und Eichreizdreieck hat ein Eichreizanteil immer negatives Vorzeichen. Bei der *CIE-Normfarbtafel* (Taf. 34) sind die reellen Eichreize so auf virtuelle Normalreize umgerechnet, daß der Spektralfarbenzug ganz vom Normalreizdreieck umschlossen wird. Dadurch haben alle Farbreize positive Normalreizanteile. Zur besseren Übertragung der Farborte aus der Farbtafel auf Millimeterpapier sind die Normalreizanteile x und y als rechtwinklige Koordinaten eingetragen. Der Weißpunkt liegt bei $x = y = z = 1/3$. Den Normalreizen X und Z ist der Helligkeitsbeiwert Null, dem Normalreiz Y der Helligkeitsbeiwert Eins zugeordnet. Damit liegen dem Auge gleich hell erscheinende Farbtöne auf einer Geraden y = konst., sie haben den gleichen Normalreizanteil y. Alle Farben des gleichen Farbtones liegen auf der Geraden, die den Weißpunkt

und die Farbe enthält. Sie enthält auch den Spektralreiz desselben Farbtones. Seine Wellenlänge ist die die Farbe kennzeichnende farbtongleiche Wellenlänge. Bei den Purpurfarben wird die kompensative Wellenlänge mit einem negativen Vorzeichen angegeben, da die Purpurfarben keine farbtongleichen Spektralreize haben. Der kompensative Spektralreiz liegt als Schnittpunkt mit dem Spektralfarbenzug auf der über den Weißpunkt hinaus verlängerten Geraden. Die Sättigung einer Farbe ist das Verhältnis ihres Abstandes vom Weißpunkt zum Abstand ihres farbtongleichen Spektralreizes vom Weißpunkt, bei den Purpurfarben das Verhältnis ihres Abstandes vom Weißpunkt zum Abstand der farbtongleichen Purpurfarben auf der Purpurgeraden zum Weißpunkt. Da die Farbtafel nur die Farbtöne und den Weißanteil oder die Sättigung anzugeben erlaubt, muß die Helligkeit z. B. als dritte Helmholtz-Koordinate neben Farbton und Sättigung bei Körperfarben durch ihren Remissions- bzw. Durchlaßgrad, bei farbigen Lichtern durch ihre Leuchtdichte bestimmt werden.

2.5.1.5. Temperaturstrahlung. Die Strahlung, die ein Körper aufgrund der ihm innewohnenden Wärmebewegung abstrahlt, heißt Temperaturstrahlung. Sie enthält elektromagnetische Strahlung aller Wellenlängen und stellt ein kontinuierliches Spektrum dar. (Vgl. Hohlraumstrahlung 2.6.10.)

Der sichtbare Anteil der Temperaturstrahlung stellt einen Farbreiz dar, dessen spektrale Zusammensetzung sich mit der Temperatur ändert. Als Farbtemperatur eines Strahlers wird die Temperatur bezeichnet, die ein schwarzer Strahler haben müßte, um den gleichen Farbreiz hervorzurufen. Bei einem schwarzen Strahler stimmen wahre Temperatur und Farbtemperatur überein. Mit der Farbtemperatur werden insbesondere als Temperaturstrahler wirkende technische Lichtquellen und die sichtbaren Sterne gekennzeichnet. Hat das Absorptionsvermögen eines Temperaturstrahlers für alle Wellenlängen denselben Wert, handelt es sich um einen grauen, andernfalls um einen selektiven Strahler. Als *schwarze Temperatur* eines Strahlers wird die Temperatur bezeichnet, die ein schwarzer Körper haben müßte, um in demselben Spektralbereich die gleiche Strahlungsdichte zu erzeugen. Die schwarze Temperatur eines Körpers ist immer niedriger als seine wahre. So beträgt sie im Krater der Anode einer Kohlebogenlampe 3500 K, während die wahre Temperatur bei 4000 K liegt. Im Strahlungsdichtediagramm schließt die Isotherme eines schwarzen Strahlers die entsprechende Isotherme eines anderen Temperaturstrahlers stets ein.

2.5.2. Die Ausbreitung des Lichtes

Die Ausbreitung des Lichtes erfolgt im Vakuum und in homogenen Medien geradlinig und allseitig. Sie wird nicht durch Licht gestört, das den Lichtweg zum Beobachter kreuzt, d. h., die Überlagerung von Licht erfolgt in ungestörter Superposition. Der von Licht sowohl in homogenen als auch in inhomogenen Medien zurückgelegte Weg läßt sich durch Anwendung des *Fermatschen Prinzips* voraussagen. Es lautet in einer für die geometrische Optik zugeschnittenen Form: Die optische Weglänge zwischen Anfangs- und Endpunkt eines Lichtweges ist im Vergleich zu den optischen Weglängen benachbarter Lichtwege extremal. Sie ist die Wegstrecke, die ein Lichtstrahl in der gleichen Zeit im Vakuum zurücklegen würde und berechnet sich als das Produkt aus Brechungsindex (vgl. 2.5.2.1.) und geometrischer Weglänge. Für inhomogene Medien ist sie gleich dem Wegintegral über den längs des Integrationsweges gültigen Brechungsindex, $\int n \, ds$. Aus dem *Fermatschen Prinzip* folgt, daß ein Lichtweg auch in der umgekehrten Richtung durchlaufen wird, wenn man Anfangs- und Endpunkt des Lichtweges vertauscht. Die Ausbreitung des Lichtes als Welle wird durch das *Huygenssche Prinzip* als das Vorrücken einer Fläche gleicher Phase beschrieben: Die zu einer Wellenfläche benachbarte Wellenfläche kann als Hüllfläche aller Elementarwellen angesehen werden, die von der gegebenen Wellenfläche mit der jeweiligen *Phasengeschwindigkeit* ausgehen.

2.5.2.1. Lichtgeschwindigkeit. Bei einem Lichtsignal hat die Lichtwelle als elektromagnetische Welle einen Anfang und ein Ende. Sie stellt eine *Wellengruppe* dar, die sich mit der *Gruppengeschwindigkeit* fortpflanzt. Die Gruppengeschwindigkeit ist die *Signalgeschwindigkeit*, denn sie bezieht sich auf die Fortpflanzung der Energie. Bei der Ausbreitung der Wellengruppe in einem Medium wird die volle Amplitude erst nach einem Einschwingvorgang erreicht, d. h. die Frontgeschwindigkeit der Wellengruppe ist von der eigentlichen Gruppengeschwindigkeit zu unterscheiden. Sie ist unabhängig vom durchlaufenen Medium und gleich der Vakuumlichtgeschwindigkeit. Innerhalb der Wellengruppe bewegen sich die Momentanwerte der elektrischen und der magnetischen Feldstärke der Lichtwelle mit der ihnen eigenen *Phasengeschwindigkeit.* Sie hängt von der Wechselwirkung der Lichtwelle mit dem Medium im eingeschwungenen Zustand ab, so, als ob die Welle keinen Anfang und kein Ende hätte. Die periodisch sich wiederholenden Phasen der Lichtwelle sind für den Beobachter ununterscheidbar. Die Phasengeschwindigkeit kann keine Signalgeschwindigkeit sein. Sie ist aber zur Beschreibung der optischen Phänomene Brechung, Doppelbrechung, Dispersion u. a. mit Hilfe des Wellenmodells notwendig. Deshalb kommt ihr in der klassischen Optik eine besondere Bedeutung zu.

Die *Vakuumlichtgeschwindigkeit* ist eine naturkonstante Größe. Ihr Zahlenwert ist $c = 299\,792\,458 \text{ ms}^{-1}$. Sie hat keine „Rückerinnerung" an den Bewegungszustand der emittierenden Lichtquelle und ist von diesem unabhängig. Gruppen-, Phasen- und Frontgeschwindigkeit sind im Vakuum gleich. Es findet keine *Dispersion* statt. Der Quotient aus der Vakuumlichtgeschwindigkeit und der Phasengeschwindigkeit des Lichtes in einem Medium ist die *Brechzahl* oder der *Brechungsindex* $c/u = n$. Von zwei Medien wird das mit dem größeren Brechungsindex das optisch dichtere genannt. Ist die Phasengeschwindigkeit des Lichtes in allen Richtungen gleich groß, ist das Medium optisch isotrop. Für einen Beobachter ist in einem *optisch isotropen Medium* auch dann die Phasengeschwindigkeit des Lichtes in allen Richtungen gleich groß, wenn sich das Medium gleichförmig bewegt und der Beobachter an dieser Bewegung teilnimmt. Bewegt sich das lichtführende Medium auf den Beobachter mit der Relativgeschwindigkeit v zu, so erhöht sich die Phasengeschwindigkeit des Lichtes für den Beobachter durch *Mitführung* um den Betrag $v(1 - 1/n^2)$. Der Faktor $(1 - 1/n^2)$ wird der *Fresnelsche Mitführungskoeffizient* genannt. Entsprechend erniedrigt sich die Phasengeschwindigkeit des Lichtes um den gleichen Betrag, wenn sich das Medium mit der Relativgeschwindigkeit vom Beobachter wegbewegt. Die Phasengeschwindigkeit des Lichtes in einem Medium ist von seiner Frequenz abhängig. Entsprechend ist der Brechungsindex eine Funktion der Frequenz

oder – wegen des Zusammenhanges $f \cdot \lambda = c$ – formal auch als eine Funktion der Vakuumwellenlänge des Lichtes schreibbar. Die Änderung des Brechungsindex mit der Vakuumwellenlänge $dn/d\lambda$ ist die *Dispersion*.

In einem *optisch anisotropen Medium* breitet sich eine Lichtwelle in Form zweier linear polarisierter Teilwellen mit unterschiedlicher Phasengeschwindigkeit aus. Die entsprechenden Brechungsindizes sind Funktionen der Hauptbrechungsindizes n_X, n_Y, n_Z und der Richtung der Wellennormalen. Die Schwingungsebenen der beiden Teilwellen stehen immer senkrecht aufeinander (vgl. 2.5.2.3.).

In einem *optisch aktiven Medium* breitet sich eine Lichtwelle in Form zweier zirkularpolarisierter Teilwellen mit entgegengesetztem Windungssinn und unterschiedlicher Phasengeschwindigkeit aus. Die entsprechenden Indizes werden mit n_+ und n_- bzw. mit n_R und n_L bezeichnet. Ihr arithmetisches Mittel ist bei *optisch einachsigen Kristallen* der Grenzwert, dem der Brechungsindex für die einfache Doppelbrechung bei Annäherung der Wellennormalenrichtung an die Richtung der *optischen Achse* zustrebt. Bei Lösungen ist ihr arithmetisches Mittel der Brechungsindex des razemischen Gemisches gleicher Konzentration des gelösten optisch aktiven Stoffes.

2.5.2.2. Absorption, Reflexion, Brechung, Dispersion. Beim Durchgang durch ein Medium wird ein Lichtstrom Φ i. allg. durch Absorption und durch Streuung geschwächt. Bei der Absorption wird durch Wechselwirkung mit dem Medium ein Teil der Strahlungsenergie in eine andere Energieform umgewandelt. Bei der Streuung wird durch Wechselwirkung mit dem Medium ein Teil des Lichtstromes aus der ursprünglichen Richtung abgelenkt. Für sehr kleine Wegstrecken Δd in einem Medium ist die Abnahme des Lichtstromes $-\Delta\Phi$ der durchlaufenen Strecke und dem Lichtstrom proportional, $-\Delta\Phi = \mu\Phi\Delta d$. Der Proportionalitätsfaktor μ ist der *Extinktionskoeffizient*. Er setzt sich aus dem Absorptionskoeffizienten τ und dem Streukoeffizienten σ zusammen, $\mu = \tau + \sigma$. Für endliche Durchlaufstrecken d in einem homogenen Medium gilt dann für einen parallel gebündelten Lichtstrom $\Phi_2 = \Phi_1 \exp(-\mu d)$. Mißt man einfallenden (Φ_1) und austretenden Lichtstrom (Φ_2) und den Weg d, dann bestimmt sich der Extinktionskoeffizient zu $\mu = (1/d) \ln(\Phi_1/\Phi_2)$. Die zur Extinktion führende Wechselwirkung des Lichtes mit dem Medium ist wellenlängenabhängig, aus anfänglichem weißem Licht wird farbiges Licht (vgl. 2.5.1.3.). Mit Eindringtiefe bezeichnet man die in einem Medium von Licht zurückgelegte Strecke, bei der durch Extinktion der Lichtstrom auf den e-ten Teil des Anfangswertes geschwächt ist. Die *Eindringtiefe* ist bei homogenen Medien gleich dem Kehrwert des Extinktionskoeffizienten. Im gesamten Spektrum der elektromagnetischen Wellen gibt es nur drei Wellenlängenbereiche, bei denen die Eindringtiefe der aus dem Weltall in die Erdatmosphäre eindringenden Strahlung bis zur Erdoberfläche reicht. Es sind dies die sogenannten Fenster: sichtbares Licht, ein Wellenlängenbereich bei $\lambda = 10\,\mu$m, in dem die Maxima der Ausstrahlung der Erdoberfläche und der Eigenstrahlung der Atmosphäre entsprechend einer Temperatur von 300 K liegen, und ein Wellenlängenbereich von $\lambda = 1$ cm bis mindestens 20 m, wobei die obere Grenze je nach Sonnenstand und -aktivität durch die Leitfähigkeit der Ionosphäre festgelegt ist.

Reflexion. An jeder Grenzfläche zwischen zwei Medien mit unterschiedlichem Brechungsindex wird entweder ein Teil oder bei Totalreflexion der gesamte auf die Grenzfläche auftreffende Lichtstrom reflektiert. Das Verhältnis von reflektiertem zu auffallendem Lichtstrom ist das Reflexionsvermögen r der Grenzfläche. Entsprechend ist die Durchlässigkeit der Grenzfläche das Verhältnis aus durchgelassenem zu einfallendem Lichtstrom. Aus Gründen der Energieerhaltung ist die Summe aus Reflexionsvermögen und Durchlässigkeit gleich Eins. Je nach geometrischer Beschaffenheit der Grenzfläche erfolgt die Reflexion diffus oder spiegelnd. Zwischen dem Lot auf die Grenzfläche und einfallendem Lichtstrahl liegt der *Einfallswinkel*. Der gespiegelte Lichtstrahl schließt mit dem Lot den *Reflexionswinkel* ein. Beide sind gleich. Einfallender Strahl, Lot und gespiegelter Strahl liegen in einer Ebene, der *Einfallsebene*. Das Reflexionsgesetz von der Gleichheit von Einfalls- und Reflexionswinkel folgt aus dem *Fermatschen Prinzip* und kann mit dem *Huygensschen Prinzip* erklärt werden (vgl. 2.5.2.; Abb. 2.5.2–1).

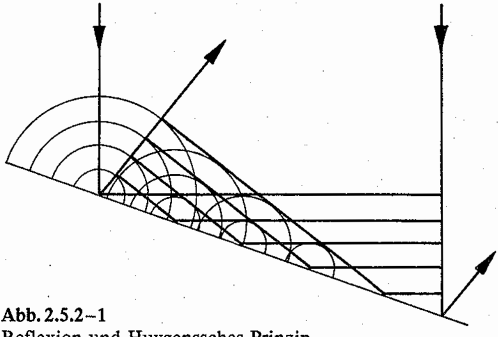


Abb. 2.5.2–1
Reflexion und Huygenssches Prinzip

Das Reflexionsvermögen einer ebenen Grenzfläche zwischen zwei durchsichtigen Medien ist bei senkrechtem Einfall und Vernachlässigung der Absorption gleich dem Quadrat aus dem Quotienten von Differenz und Summe der Brechungsindizes der beiden Medien, $r = (n_2 - n_1)^2/(n_2 + n_1)^2$. Dabei ist es gleichgültig, auf welcher Seite der Grenzfläche das Licht einfällt. Liegt zwischen den beiden Medien eine Grenzschicht, deren Dicke größer als die Wellenlänge des auf sie auftreffenden Lichtes ist und in der sich der Brechungsindex kontinuierlich von dem einen auf den anderen Wert ändert, ist die Reflexion bei senkrechtem Einfall nahezu aufgehoben. Bei einer Dicke $d < \lambda/4$ reflektiert eine Grenzschicht praktisch wie eine Grenzfläche. Beim Auftreffen von Licht auf ein undurchsichtiges bzw. stark absorbierendes Medium hat man zur Darstellung des Reflexionsvermögens mit dessen Brechungsindex auch seinen Absorptionsindex $\varkappa$ zu berücksichtigen, $n' = n(1 + i\varkappa)$. Bei senkrechtem Einfall gilt wie im Fall ohne Absorption für das Reflexionsvermögen das Quadrat von Differenz und Summe der Brechungsindizes als Quotient,
$r = [(n_2 - n_1)^2 + n_2^2\varkappa_2^2]/[(n_2 + n_1)^2 + n_2^2\varkappa_2^2]$, wobei Medium 1 nichtabsorbierend sein soll. Bei der Reflexion kommt es i. allg. durch Wechselwirkung mit dem Medium hinter der Grenzfläche zur Änderung des Polarisationszustandes. Bei unpolarisiert einfallendem Licht ist nach Reflexion das Licht teilweise oder ganz, bei hinter der Grenzfläche nichtabsorbierendem Medium (Ausnahme bei Totalreflexion) linear, bei absorbierendem Medium elliptisch polarisiert. Bei senkrechtem Einfall und Reflexion am optisch dichteren, nichtabsorbierenden Medium findet ein Phasensprung um π bzw. $\lambda/2$ statt, bei Reflexion am optisch dünneren Medium erfolgt die Reflexion ohne Phasensprung. Für den Fall der Reflexion an einem absorbierenden Medium 2 gilt

für den Phasensprung $\tan \delta_r = 2n_2 \varkappa_2/(n_1^2 - n_2^2 - n_2^2 \varkappa_2^2)$, Medium 1 ist nicht absorbierend. Im Durchlicht klare Stoffe kann man näherungsweise stets als nichtabsorbierend ansehen, für Luft kann man den Brechungsindex näherungsweise Eins setzen.

Brechung. Ein durch die Grenzfläche zwischen zwei optisch isotropen Medien als Lichtstrahl hindurchtretender Lichtstrom wird gebrochen. Einfalls- und Brechungswinkel werden vom Lot auf die Grenzfläche aus gerechnet. Das Brechungsgesetz lautet: Die Sinus von Einfalls- und Brechungswinkel verhalten sich zueinander wie die Phasengeschwindigkeiten in den beiden Medien und damit umgekehrt zueinander wie deren Brechungsindizes, $\sin \alpha_1/\sin \alpha_2 = u_1/u_2 = n_2/n_1$. Das so formulierte Brechungsgesetz läßt sich wie das Reflexionsgesetz mit dem Huygensschen Prinzip erklären (Abb. 2.5.2-2) und steht im Einklang mit dem Fermatschen Prinzip. Stehen reflektierter und durch die Grenzfläche hindurchtretender Anteil des Lichtstromes als Lichtstrahlen senkrecht aufeinander, ist der reflektierte Anteil vollständig linear polarisiert. Seine Schwingungsebene steht senkrecht zur Einfallsebene. Der zugehörige Einfallswinkel heißt *Brewsterscher Winkel*, das Gesetz *Brewstersches Gesetz.*

Totalreflexion. Findet die Brechung beim Übergang vom optisch dichteren zum optisch dünneren Medium statt, ist der Brechungswinkel stets größer als der Einfallswinkel. Das Brechungsgesetz kann nur für Brechungswinkel $\leqq 90°$ erfüllt werden. Überschreitet der Einfallswinkel den nach dem Brechungsgesetz zu einem Brechungswinkel von 90° zugehörigen Wert, kommt es zur *Totalreflexion.* Das Licht ist nach der Totalreflexion i. allg. partiell elliptisch polarisiert.

Dispersion. Wird ein Lichtstrahl an der Grenzfläche zwischen zwei Medien gebrochen und weist eines der beiden

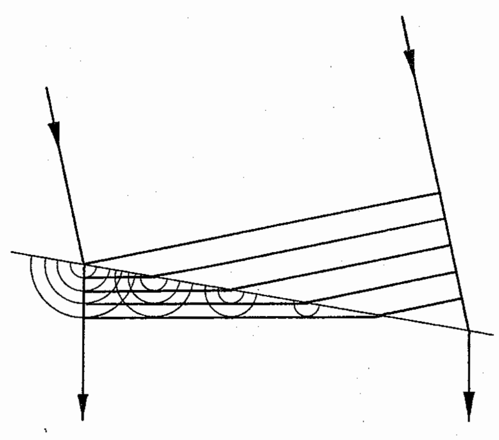


Abb. 2.5.2-2 Brechung und Huygenssches Prinzip

Tab. 2.5.2-3 Spezifisches Drehvermögen in °/mm optisch aktiver Kristalle

	Fraunhofersche Linie B	D	E	F	H
Quarz	15,75	21,71	27,54	32,76	51,19
$NaClO_3$	2,27	3,13	3,74	4,67	7,17

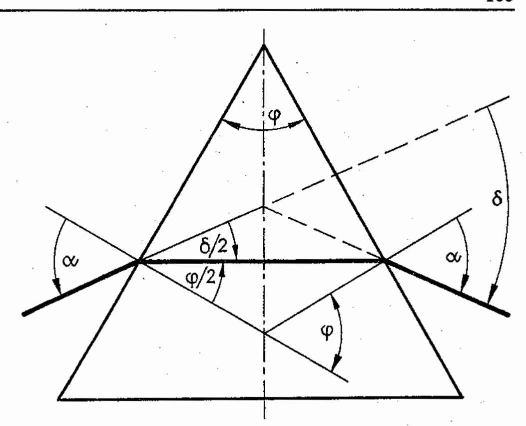


Abb. 2.5.2-4 Strahlengang durch ein optisches Prisma im Minimum der Ablenkung

Medien eine Dispersion auf, kommt es zu einer Zerlegung des Lichtes in sein Spektrum. Weil beim Übergang von Luft auf Glas das blaue Licht stärker als das rote gebrochen wird und man anfangs nur diesen Fall kannte, nennt man die Dispersion, bei der der Brechungsindex mit abnehmender Wellenlänge zunimmt, normal ($dn/d\lambda < 0$), im umgekehrten Fall anomal ($dn/d\lambda > 0$). Entsprechend nennt man eine *Rotationsdispersion* normal, wenn der Drehwinkel der Schwingungsebene von linear polarisiertem Licht beim Durchgang durch ein optisch aktives Medium mit abnehmender Wellenlänge zunimmt, anomal, wenn er mit abnehmender Wellenlänge abnimmt. Als Maß der Dispersion wird die Differenz der Brechungsindizes eines Mediums für zwei *Frauenhofersche Linien* benutzt: $n_H - n_C$ ist die spezifische Dispersion, $n_F - n_C$ die mittlere Dispersion, $\nu = (n_D - 1)/(n_F - n_C)$ *Abbesche Zahl.* Die Indizes C, D, F und H beziehen den Brechungsindex n auf die entsprechenden Fraunhoferschen Linien. Ein Maß für die Rotationsdispersion in Analogie zu dem für die Dispersion gibt es nicht. Bei festen Stoffen wird zur Charakterisierung der Rotationsdispersion die spezifische Drehung (α) oder das spezifische Drehvermögen, das ist der auf eine Weglänge von 1 mm bezogene Drehwinkel für eine bestimmte Wellenlänge, angegeben (Tab. 2.5.2-3). Bei Flüssigkeiten und Lösungen optisch aktiver Stoffe in inaktiven Lösungsmitteln gibt man eine spezifische Drehung (α) an, indem man den Drehwinkel auf einen Lichtweg von 1 dm und bei Lösungen zusätzlich auf eine Konzentration der gelösten Stoffmenge von 1 g je 1 cm³ Lösung bezieht, (α) hat dann die Einheit $(° \cdot cm^3)/(dm \cdot g)$.

Bei schiefem Durchgang durch ein von zwei planparallelen Grenzflächen eingeschlossenes homogenes Medium erfährt ein Lichtstrahl und bei Dispersion jeder Anteil des Lichtstrahles mit unterschiedlicher Wellenlänge für sich nur eine Parallelverschiebung, sofern Brechungsindex und Dispersion des Mediums vor und hinter dem von den Grenzflächen eingeschlossenen Medium gleich sind.

Ein *optisches Prisma* wird verkörpert durch ein homogenes prismatisch begrenztes Medium. Zwei der ebenen Begrenzungsflächen sind die brechenden Flächen. Sie schließen den brechenden Winkel φ ein und schneiden sich in der brechenden Kante. Dem brechenden Winkel φ gegenüber liegt die Basis als dritte Prismenfläche. Jede senkrecht zur brechenden Kante liegende Ebene ist ein Hauptschnitt des

Prismas. Bei symmetrischem Strahlengang im *Hauptschnitt* ist die Ablenkung δ des Lichtstrahles ein Minimum (Abb. 2.5.2–4).
Kann für den Außenraum des Prismas der Brechungsindex gleich Eins gesetzt werden, so folgt aus dem Brechungsgesetz der Brechungsindex des Prismenmediums zu $n = \sin[(\delta + \varphi)/2]/\sin\varphi/2 = \sin\alpha/\sin\varphi/2$ wegen $\beta = \varphi/2$ und $\alpha = (\delta + \varphi)/2$. Ist der brechende Winkel φ sehr klein, dann ist auch die Ablenkung sehr klein. Wegen $\sin\varphi \approx \varphi$ folgt für die Ablenkung $\delta = (n-1)\varphi$. Werden zwei optische Prismen, deren brechende Winkel sich umgekehrt zueinander verhalten wie ihre Dispersionen, so im Strahlengang verwendet, daß die beiden brechenden Kanten parallel jedoch so liegen, daß die eine bei der Basis des anderen Prismas liegt und umgekehrt, kommt es beim Durchgang eines Lichtstrahles im Hauptschnitt zu einer Ablenkung ohne Zerlegung des Lichtes in sein Spektrum (Abb. 2.5.2–5). Ein solches aus zwei Teilprismen bestehendes optisches Prisma wird *achromatisch* genannt.

2.5.2.3. Doppelbrechung, Interferenz, Kohärenz. Unter *Doppelbrechung* versteht man gewöhnlich die Aufspaltung eines Lichtstrahles in zwei Teilstrahlen durch Brechung in Anlehnung an die Erfahrung, daß man z. B. die Schrift auf einem Blatt Papier durch ein daraufliegendes Kalkspatstück hindurch doppelt sieht. Bei senkrechtem Auftreffen eines Lichtstrahles auf die Spaltfläche eines Kalkspatrhomboeders setzt nur einer der beiden Teilstrahlen seinen Weg geradlinig fort. Nur er durchläuft den Kristall, als ob dieser eine planparallele Glasplatte wäre. Deshalb wird er der ordentliche Strahl genannt. Der andere, außerordentliche Teilstrahl wird beim Einfallswinkel Null gebrochen. Er verläßt den Kristall parallel verschoben (Abb. 2.5.2–6).
Im strengen Sinne versteht man unter Doppelbrechung die Aufspaltung einer monochromatischen Lichtwelle in zwei Teilwellen mit unterschiedlicher Phasengeschwindigkeit. Das Maß der Doppelbrechung ist die Differenz der Brechungsindizes für die beiden Teilwellen. In diesem Sinne ist die Aufspaltung von Lichtwellen in jeweils zwei gegenläufig zirkular polarisierte Teilwellen mit unterschiedlicher Phasengeschwindigkeit durch ein optisch aktives Medium auch eine Doppelbrechung. Sie läßt sich durch Brechung im eigentlichen Sinne nachweisen, wenn man ein Quarzdoppelprisma nach Fresnel benutzt (Abb. 2.5.2–7).
Werden beim Eintritt von linear polarisiertem Licht in das optisch aktive Medium die paarweise erzeugten und gegenläufig zirkular polarisierten Teilwellen nicht durch Brechung räumlich getrennt, überlagern sie sich nach Verlassen des optisch aktiven Mediums wieder zu linear polarisiertem Licht. Die für alle Teilwellenpaare einer Wellenlänge gleiche Phasendifferenz führt zu einer Drehung der Schwingungsebene, die für Licht verschiedener Wellenlänge verschieden groß ist und *Rotationsdispersion* genannt wird. Ein optisch aktives Medium ist rechts- oder linksdrehend, je nachdem, wie für den Beobachter die Schwingungsebene des ihm entgegenkommenden Lichtes gedreht wird. Die Drehung wird rückgängig gemacht, wenn das Licht nach Spiegelung das optisch aktive Medium auf demselben Wege in umgekehrter Richtung durchläuft, im Gegensatz zur Drehung in einem durch den *Faraday-Effekt* optisch aktiv gemachten Medium. In einem solchen wird der Drehwinkel auf dem Rückweg um den gleichen Betrag erhöht. Beim Faraday-Effekt ist der Drehwinkel dem parallel

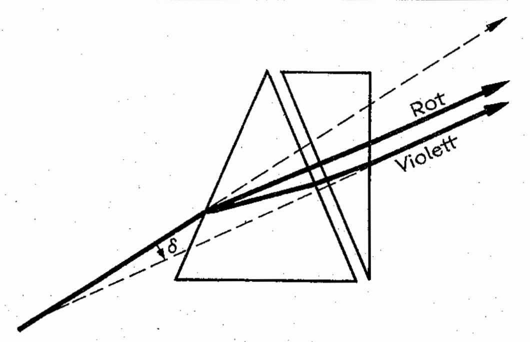


Abb. 2.5.2–5 Strahlengang durch ein achromatisches Prisma

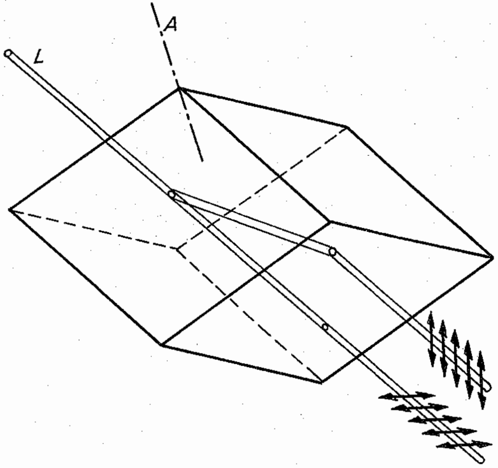


Abb. 2.5.2–6 Strahlengang durch ein Kalkspat-Spaltrhomboeder. Einfallender Lichtstrahl L unpolarisiert, nach Verlassen des Kristalles ist jeder Teilstrahl linear polarisiert, A ist die optische Achse des Kristalles

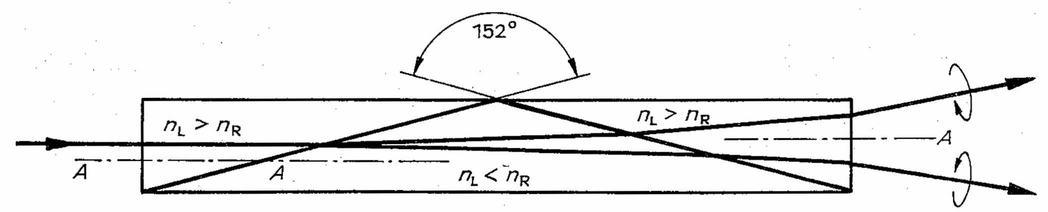


Abb. 2.5.2–7 Strahlengang in einem Quarzdoppelprisma nach Fresnel. Die optischen Achsen A der Teilprismen sind gleichorientiert

zur Ausbreitungsrichtung des Lichtes wirksamen Magnetfeld H und der Länge l des Lichtweges proportional, $\alpha = VlH$. Der Proportionalitätsfaktor V, die Verdetsche Konstante, ist eine Stoffkonstante. Da sie von der Wellenlänge des Lichtes abhängig ist, tritt auch beim Faraday-Effekt eine Rotationsdispersion auf. Während bei der strukturbedingten optischen Aktivität auf Hin- und Rückweg der Drehsinn erhalten bleibt, kehrt er bei der mit einem Magnetfeld erzeugten optischen Aktivität mit dem Magnetfeld sein „Vorzeichen“ um.

Die *optische Aktivität* beruht auf der Wechselwirkung der Lichtwellen entweder mit schraubenförmigen Strukturen der Atomanordnung im Kristallgitter, z. B. beim Quarz, oder mit den durch ihre Asymmetrie als optisch aktive Moleküle erkennbaren Bausteinen eines Kristalles, z. B. beim Zucker. Bei Flüssigkeiten und Lösungen optisch aktiver Stoffe in optisch inaktiven Lösungsmitteln beruht die optische Aktivität ebenso wie bei den Gasen auf der Wechselwirkung des Lichtes mit den asymmetrisch strukturierten Molekülen. Enthält eine Lösung optisch aktive spiegelbildlich isomere Moleküle in gleicher Konzentration, kristallisiert aus ihr ein Razemat aus. Es ist – wie bereits vorher die Lösung – optisch inaktiv. Von sich aus optisch aktive Gase, Flüssigkeiten und Lösungen sind dennoch als optisch isotrop zu bezeichnen. Das trifft auch für optisch aktive Kristalle des regulären oder kubischen Kristallsystemes zu, z. B. für Natriumchlorat $NaClO_3$. Eine bei optisch einachsigen Kristallen auftretende optische Aktivität ist immer auf einen kleinen Winkelbereich der Ausbreitungsrichtung des Lichtes symmetrisch zur optischen Achse beschränkt. Sofern keine optische Aktivität vorliegt, entstehen bei der Doppelbrechung zwei linear polarisierte Teilwellen. Ihre Schwingungsebenen stehen aufeinander senkrecht. Während beim Eintritt in das optisch aktive Medium jede linear polarisierte Lichtwelle in die beiden gegenläufig zirkular polarisierten Teilwellen aufspaltet, muß bei der Doppelbrechung für das Zustandekommen einer linear polarisierten Teilwelle im Licht vor dem Übergang in das doppelbrechende Medium diese Schwingungsrichtung durch einen **E**-Vektor oder durch eine Komponente des **E**-Vektors in dieser Richtung vertreten sein! Bis auf wenige als Grenzfälle anzusehende Ausnahmen haben die Wellennormalen der beiden Teilwellen durch die Doppelbrechung auch eine unterschiedliche Richtung. Die mit der Doppelbrechung verursachte Aufspaltung eines Lichtstrahles in zwei Teilstrahlen als parallel begrenzte Lichtbündel ist davon zu unterscheiden. Strahlrichtung und Wellennormalenrichtung sind nach Doppelbrechung i. allg. verschieden (Abb. 2.5.2–8). Allgemein gilt, daß ein Medium doppelbrechend ist, wenn es optisch anisotrop ist. Die optische Anisotropie kann verschiedene Ursachen haben.

1. *Kerr-Effekt:* in Gasen und Flüssigkeiten Gleichorientierung von Molekülen mit einem elektrischen Dipolmoment in Richtung eines äußeren elektrischen Feldes;
2. *Cotton-Mouton-Effekt:* in Gasen und Flüssigkeiten Gleichorientierung von Molekülen mit einem magnetischen Moment in Richtung eines äußeren magnetischen Feldes;
3. *Strömungsdoppelbrechung:* in Flüssigkeiten Gleichorientierung nichtkugelsymmetrischer Moleküle durch laminare Scherströmung;
4. *Anisotropie der Kristallstruktur;*
5. *Spannungsdoppelbrechung:* in ursprünglich isotropen Feststoffen elastische Deformation durch Scherung, einseitigen Zug bzw. Druck oder innere Spannungen;

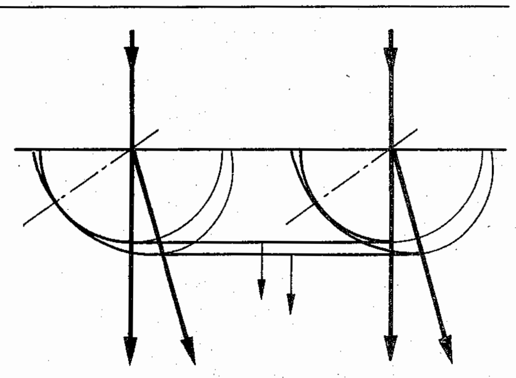


Abb. 2.5.2–8 Doppelbrechung und Huygenssches Prinzip

6. in Dielektrika spontane oder induzierte elektrische Polarisation;
7. in kristallinen Flüssigkeiten gebietsweise Parallelanordnung der Längsachsen der Moleküle.

Da die Phasengeschwindigkeit des Lichtes in einem Medium von dessen Dielektrizitätskonstante abhängt, wird der Brechungsindex in optisch anisotropen Medien der Schwingungsrichtung des **D**-Vektors der Lichtwelle zugeordnet. Denkt man sich in allen möglichen Schwingungsrichtungen die zugehörigen Brechungsindizes von einem Punkt aus als Strecken abgetragen, bilden ihre Endpunkte eine geschlossene Fläche zweiten Grades, die nach Fletcher *Indikatrix*, gelegentlich auch *Cauchysches Polarisationsellipsoid* genannt wird. Das optische Verhalten eines doppelbrechenden Mediums, insbesondere das der Kristalle, wird besonders anschaulich und übersichtlich, wenn man den Betrachtungen die Indikatrix zugrunde legt. Im Fall optisch isotroper Medien und der Kristalle im regulären oder kubischen *Kristallsystem* ist die Indikatrix eine Kugel. Bei Kristallen im trigonalen, tetragonalen und hexagonalen Kristallsystem ist sie ein Rotationsellipsoid. Senkrecht zur Rotationsachse schneidet eine durch die Mitte der Indikatrix gelegte Ebene, eine Diametralebene, die Indikatrix in einem Kreis. Die damit festgelegten Brechungsindizes sind alle gleich groß und damit unabhängig von der Schwingungsrichtung der Lichtwellen, die sich in Richtung der Rotationsachse ausbreiten. Es findet keine Doppelbrechung statt. Die Rotationsachse der Indikatrix ist optische Achse, sie fällt stets mit der trigonalen, tetragonalen bzw. hexagonalen Drehachse des entsprechenden Kristallsystems und damit mit einer fest vorgegebenen Richtung in einem entsprechenden Kristall zusammen. Solche Kristalle sind optisch einachsig. Bei Kristallen im orthorhombischen, monoklinen und triklinen Kristallsystem und i. allg. Fall ist die Indikatrix ein dreiachsiges Ellipsoid mit drei aufeinander senkrecht stehenden Hauptachsen X, Y, Z. Von allen möglichen Diametralebenen schneiden zwei die Indikatrix in einem Kreis, alle anderen in einer Ellipse. Die Richtungen senkrecht zu den beiden Kreisschnitten sind wiederum optische Achsen. Kristalle in diesen Kristallsystemen sind optisch zweiachsig. Ist senkrecht zu einer gegebenen Ausbreitungsrichtung des Lichtes der Diametralschnitt mit der Indikatrix eine Ellipse, werden in der Natur von allen Schwingungsrichtungen nur die in Richtung der großen und der kleinen Halbachse der Ellipse zugelassen. Das ist gleichbedeutend damit, daß nur zwei Teilwellen zustande kommen können, deren Phasengeschwindigkeit bei der einen die kleinstmög-

liche und bei der anderen die größtmögliche für diese Ausbreitungsrichtung ist. Die Natur befolgt hier ein Extremalprinzip. Ist die Indikatrix ein Rotationsellipsoid, enthält jeder Diametralschnitt als eine Ellipsenhalbachse den Kreisschnittradius, für eine der beiden Wellen ist somit der Brechungsindex unabhängig von der Ausbreitungsrichtung stets derselbe. Er wird der Brechungsindex des ordentlichen Strahles, der andere der des außerordentlichen Strahles genannt und mit n_0 bzw. n_e bezeichnet. Bei allen doppelbrechenden Medien ergeben sich von einem Punkt ausgehend entsprechend den beiden extremalen Phasengeschwindigkeiten mit aufeinander senkrecht stehenden Schwingungsrichtungen der beiden Wellen für jede Ausbreitungsrichtung über alle Richtungen zusammengefaßt zwei Wellenflächen als Flächen gleicher Phase, wie sie die von einem Punkt ausgehenden synchron schwingenden Elementarwellen aufweisen. Das Zustandekommen der Doppelbrechung läßt sich somit genauso mit Hilfe des Huygensschen Prinzips veranschaulichen wie die einfache Brechung beim Übergang von Licht in ein optisch isotropes Medium (Abb. 2.5.2–8).

Interferenz. Hierunter versteht man die Gesamtheit von Erscheinungen, die durch Überlagerung von zwei Wellen an einem Ort hervorgerufen werden. Alle beobachtbaren Interferenzerscheinungen des Lichtes von spontan emittierenden Lichtquellen entstehen dadurch, daß sich jeweils zwei aus einer Welle hervorgegangene Teilwellen im Beobachtungsraum überlagern. Es ist allerdings gelungen, mit dem Licht von zwei moden- und phasenstabilisierten Lasern Interferenzerscheinungen zu erzeugen. Dieser Fall entspricht der Erzeugung von Interferenzerscheinungen mit elektromagnetischen Rundfunkwellen von zwei Sendern. Das Interferieren von Lichtwellen mit sich selbst zu stehenden Wellen nach einer spiegelnden Reflexion ist ein spezieller Fall von Interferenz. Durch Überlagerung der spontan emittierten Lichtwellen von zwei Punkten einer Lichtquelle oder durch Überlagerung der Lichtwellen von zwei getrennten Lichtquellen kann man keine beobachtbaren Interferenzerscheinungen erzeugen, da es sich bei solchen Lichtwellen um eine Folge von Wellengruppen mit stets wechselnder Phase und Polarisation handelt. Deren momentane Interferenzen sind nicht beobachtbar, weil sie in einem viel zu schnellen Wechsel zeitlich und räumlich erfolgen. Soll eine optische Interferenzerscheinung beobachtbar sein, müssen alle im Verlauf der Beobachtungszeit daran beteiligten Lichtwellen an einem Beobachtungsort paarweise durch Überlagerung dieselbe Interferenzerscheinung hervorrufen. Die zu einem Paar gehörenden Wellen müssen kohärent sein.

Kohärenz. Zwei Lichtwellen sind kohärent, wenn sie sich trotz ihrer endlichen Länge als Wellengruppen am Beobachtungsort überlagern und ihre Wellenlänge und Phasendifferenz dabei hinreichend konstant sind. Eine Forderung nach Kohärenz schließt ein, daß der Gangunterschied der sich überlagernden Wellengruppen als Differenz der optischen Weglängen nicht größer als die Kohärenzlänge sein darf. Als Maß für die Kohärenzlänge hat sich der Quotient $\lambda^2/2\delta$ mit δ als Halbwertsbreite der in der Wellengruppe enthaltenen Wellenlängenverteilung bewährt. Die Kohärenzlänge von Lichtwellen einer Spektrallinie liegt in der Größenordnung von Metern. Die sogenannte *Kohärenzbedingung* $d \sin\alpha \ll \lambda/2$ besagt, daß eine mit spontan emittierender Lichtquelle erzeugte Interferenzerscheinung nur beobachtbar ist, wenn das Produkt aus seitlicher Ausdehnung d der Lichtquelle und dem Sinus der Beleuchtungsapertur der Interferenzanordnung klein gegen die halbe Wellenlänge des Lichtes ist. Sie beinhaltet die Forderung, daß die Interferenzfiguren, die das von zwei verschiedenen Punkten einer Lichtquelle ausgehende Licht auf einem Bildschirm erzeugt, nur wenig gegeneinander verschoben sein dürfen, daß die maximal zulässige Verschiebung klein sein muß gegen den Abstand zwischen einem Intensitätsmaximum und dem darauf folgenden Intensitätsminimum der Interferenzfigur.

2.5.2.4. Beugung und Streuung. Unter Beugung versteht man Vorgänge, die zur Abweichung von dem nach der geometrischen Optik erwarteten Strahlenverlauf führen. Beugungsphänomene treten an Schattengrenzen auf, d. h. bei der seitlichen Begrenzung eines Lichtbündels durch Kanten, Spalte und körperliche Blenden. Nach dem *Babinetschen Prinzip* rufen komplementäre Hindernisse im Strahlengang die gleiche Beugungserscheinung hervor. Unter zwei komplementären Hindernissen versteht man solche, die sich zu einem geschlossenen Hindernis ergänzen, z. B. eine kreisrunde Scheibe und ein kreisrundes Loch von gleichem Durchmesser. Bei der Beugung wird die Frequenz des Lichtes durch den Beugungsvorgang nicht geändert. Auf Beugung beruht die Grenze des Auflösungsvermögens optischer Geräte. Beugung am Gitter ist Beugung an geordneten Strukturen, auch wenn die Beugung im eigentlichen Sinne nur bei der Wechselwirkung des Lichtes mit dem einzelnen Strukturelement stattfindet. Einfache Vertreter einer geordneten Struktur in diesem Sinne sind die optischen Gitter. Bei diesen überlagern sich die von den einzelnen Gitterelementen kohärent ausgehenden Elementarwellen bei einem Gangunterschied von ganzzahligen Vielfachen der Wellenlänge λ zu Intensitätsmaxima. In Richtung der Intensitätsmaxima ist die Intensität durch Beugung am Gitterelement mitbestimmt. Ein Intensitätsmaximum erscheint bei der Beugung am Gitter nicht, wenn es mit einer Nullstelle der Beugung am Gitterelement zusammenfällt. Sind Lichtquelle und Beobachter von dem Ort, an dem der Beugungsvorgang ausgelöst wird, unendlich weit entfernt bzw. arbeitet man bei Beleuchtung und Beobachtung mit parallelem Licht, so spricht man von *Fraunhoferscher*, andernfalls von *Fresnelscher Beugung*. Bei der Anwendung des *Huygensschen Prinzipes* wird das Beugungsproblem nur geometrisch und nicht die Wechselwirkung der Lichtwelle mit dem Hindernis berücksichtigt. So sagt das Huygenssche Prinzip nichts über den Polarisationszustand der gebeugten Welle aus. Das Hindernis enthält z. T. reflektierende Oberflächen, so daß bei unpolarisiert eingestrahltem Licht im gebeugten Licht die Komponente überwiegt, deren Schwingungsrichtung senkrecht zu den strahlbegrenzenden Kanten liegt.

Die Beugung von Licht an einem kleinen Objekt nennt man Streuung. Die Streuung an einer großen Anzahl kleiner, fein verteilter Objekte führt zur Zerstreuung des Lichtes. Die Lichtausbreitung erfolgt diffus. Haben die streuenden Objekte in bezug auf die Wechselwirkung mit dem Licht nur die Eigenschaft der Lichtundurchlässigkeit, so ist die Streuung eigentlich eine Beugung und rechnerisch wie eine solche zu behandeln. Man beobachtet Streuung ohne Änderung der Frequenz des Lichtes, wie beim Himmelsblau, und Streuung mit Änderung der Frequenz des Lichtes, wie beim Smekal-Raman- oder Compton-Effekt, und spricht dann von kohärenter und inkohärenter, bei der Elektronenbeugung entsprechend von elastischer und von unelastischer Streuung. Die Wechselwirkung des Lichtes mit ungeordneten kleinen Bereichen ein und desselben Me-

diums, die sich von ihrer Umgebung durch ihre optische Dichte oder durch ihren Ordnungsgrad unterscheiden, führt auch zur Streuung. Unpolarisiert eingestrahltes Licht ist nach Streuung um 90° vollständig linear polarisiert.

2.5.3. Die Abbildung

Die optische Abbildung ist ein Vorgang, bei dem von Objektpunkten ausgehendes Licht durch abbildende Elemente in reellen oder virtuellen Bildpunkten mehr oder weniger punktnah vereinigt wird. Ist die Punktnähe hinreichend klein, sagt man, das Bild des Objektpunktes ist scharf. Die Bildschärfe wird durch eine physiologisch bedingte Erkennbarkeitsgrenze des Auges subjektiv beurteilt. Zur Bildunschärfe führen Abbildungsfehler der abbildenden Elemente und die Beugungserscheinungen durch die Begrenzung des Strahlenganges. Selbst wenn die von einem Objektpunkt zunächst divergent ausgehenden Kugelwellen bei „fehlerfreier" Abbildung als Kugelwellen konzentrisch auf den Bildpunkt zulaufen, ist das Bild des Objektpunktes in einer Bildebene ein von mehreren hellen und dunklen Ringen umgebenes Beugungsscheibchen. Bei selbstleuchtenden Objekten sind die von den verschiedenen Objektpunkten ausgehenden Lichtwellen nicht kohärent. Im Bildpunkt addieren sich die Intensitäten. Im Gegensatz dazu ist das von verschiedenen Objektpunkten ausgehende Licht nichtselbstleuchtender, d. h. beleuchteter Objekte kohärent. Im Bildpunkt addieren sich die Amplituden. Ein durch optische Abbildung erzeugtes Bild ist i. allg. zweidimensional. Durch beidäugiges Sehen, d. h. stereoskopisches Sehen, erhält man von räumlich ausgedehnten Objekten auch einen räumlichen Bildeindruck, obwohl jedes Auge mit der Netzhaut nur ein zweidimensionales Bild aufnimmt. Stehen für jedes Auge getrennt wahrnehmbar von einem räumlich ausgedehnten Objekt je ein in der jeweiligen Blickrichtung hergestelltes optisches Abbild zur Verfügung, wird durch das stereoskopische Sehen ein dreidimensionaler Bildeindruck des Objektes erzeugt. Man kann auf diese Weise durch stereoskopische Betrachtung auch räumliche Eindrücke von Objekten erhalten, die einer direkten Beobachtung nicht zugänglich sind und erst durch vergrößernde optische Abbildung, z. B. mit einem Mikroskop oder mit einem Elektronenmikroskop, Objektdetails erkennen lassen. Im Normal-

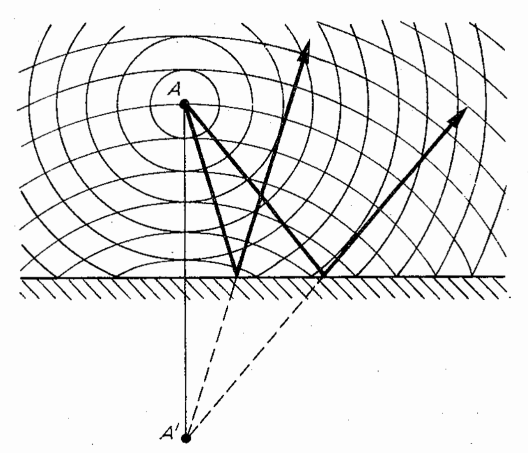


Abb. 2.5.3-1 Reflexion von Kugelwellen am ebenen Spiegel

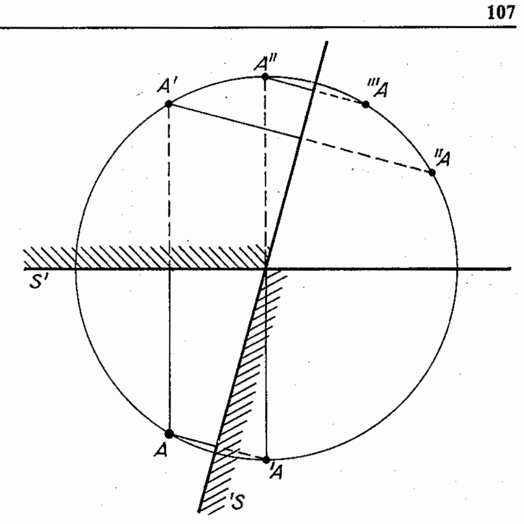


Abb. 2.5.3-2 Abbildung am Winkelspiegel $S'/'S$

fall soll eine optische Abbildung wie die durch den dioptrischen Apparat des Auges ein zentralperspektivisches Bild liefern. Aus technischen Gründen soll das durch die Abbildung erzeugte Bild möglichst in einer Ebene liegen.

2.5.3.1. Abbildung durch Reflexion. Kommt eine Abbildung durch Reflexion des Lichtes an der Grenzfläche zwischen zwei Medien zustande, hat die Grenzfläche die Funktion eines Spiegels. Sie ist das abbildende Element. *Ebene Spiegel* reflektieren die von einem Objektpunkt ausgehenden Kugelwellen so, daß sie von einem Punkt hinter der Spiegelebene auszugehen scheinen (Abb. 2.5.3-1). Dieser nur scheinbare Ausgangspunkt für das Licht ist das *virtuelle* oder *scheinbare Bild* des Objektpunktes. Es läßt sich durch rückwärtige Verlängerung der reflektierten Lichtstrahlen konstruieren und liegt auf dem Lot vom Objektpunkt auf die Spiegelfäche genau so weit hinter dem Spiegel wie der Objektpunkt vor dem Spiegel liegt. Man sieht den Objektpunkt in der geradlinigen rückwärtigen Verlängerung der in das Auge gelangenden Lichtstrahlen. Nur bei Reflexion am ebenen Spiegel hat jeder Objektpunkt unabhängig vom Öffnungswinkel des für die Abbildung zugelassenen Strahlenbündels einen Bildpunkt! *Winkelspiegel* bestehen aus zwei ebenen Spiegeln, die sich reell oder virtuell in einer Geraden, der Spiegelachse, schneiden. Jede Ebene senkrecht zur Spiegelachse ist ein Hauptschnitt. Alle virtuellen Bildpunkte eines zwischen den Spiegeln befindlichen Objektpunktes liegen mit dem Objektpunkt in demselben Hauptschnitt auf einem Kreis konzentrisch zur Spiegelachse. Die Anzahl der virtuellen Bildpunkte ist endlich, denn ein durch Spiegelung an dem einen Spiegel entstandener und hinter diesem liegender virtueller Bildpunkt wird nur dann durch Spiegelung an dem anderen Spiegel zu einem weiteren virtuellen Bildpunkt, wenn er in dem Halbraum vor dem anderen Spiegel liegt (Abb. 2.5.3-2). Zwei aufeinander senkrecht stehende, ebene Spiegel bilden den Raum, den sie mit ihren Ebenen einschließen, hinter sich wie durch eine 180°-Drehung um die Spiegelachse virtuell ab. Drei aufeinander senkrecht stehende, ebene Spiegel bilden den Raum, den sie mit ihren Ebenen einschließen, hinter sich durch eine Inversion am Schnittpunkt der drei Spiegelebenen virtuell ab. *Gekrümmte Spiegel* liefern Bildpunkte nur bei Verwendung von Strahlenbündeln mit hin-

reichend kleinem Öffnungswinkel. Da die Augenpupille auch bei kürzester Bildentfernung, der deutlichen Sehweite von 25 cm den Öffnungswinkel des für die Abbildung zugelassenen Strahlenbündels sehr stark einschränkt, sieht man Spiegelbilder in gekrümmten Spiegeln gewöhnlich auch scharf. Wendet ein Spiegel dem Licht die gewölbte Seite zu, ist es ein Wölbspiegel oder Konvexspiegel, andernfalls ein Hohlspiegel oder Konkavspiegel. *Rotationssymmetrische Spiegel* haben eine optische Achse, die Symmetrieachse. Liegt sie in der Zeichenebene, ist diese ein Meridionalschnitt. Jede Ebene, die die optische Achse enthält, ist eine Meridionalebene. Strahlen, die in einer Meridionalebene liegen, heißen Meridionalstrahlen, andernfalls windschiefe Strahlen. Den achsennahen Raum, in dem eine punktförmige ideal-geometrische Abbildung angenommen wird, nennt man den fadenförmigen oder Gaußschen Raum bzw. das *paraxiale Gebiet*. Bei punktförmiger idealgeometrischer Abbildung sind Objekt- und Bildpunkt konjugierte Punkte. Die von einem Objektpunkt ausgehenden Strahlen sind homozentrisch. Der Objektpunkt ist Konvergenzpunkt. Bei idealgeometrischer Abbildung sind auch die Bildpunktstrahlen homozentrisch, ihr Konvergenzpunkt ist der Bildpunkt, auch wenn er nur durch eine rückwärtige Verlängerung der Strahlen als virtueller Bildpunkt zustande kommt. In formaler Analogie spricht man auch von virtuellen Objektpunkten. Beispielsweise ist der geometrische Brennpunkt des Rotationshyperboloids als Wölbspiegel virtueller Objektpunkt, wenn auf ihn Lichtstrahlen als Brennpunktstrahlen zulaufen, vorher aber an dem Hyperboloid reflektiert werden und sich in dem anderen Brennpunkt des Hyperboloids als reeller Bildpunkt schneiden. So wie nach Spiegelung (oder Brechung) die rückwärtige Verlängerung der Strahlen zum virtuellen Bildpunkt führt, führt eine Vorwärtsverlängerung zum virtuellen Objektpunkt. Für brennpunktnahe Bildpunkte besteht die Notwendigkeit zur Beschränkung auf das paraxiale Gebiet nicht. So können mit einem *Parabolspiegel* beliebig großen Durchmessers achsenparallele Strahlen in einem Bildpunkt, dem Brennpunkt, vereinigt werden. Ist der Parabolspiegel Hohlspiegel, ist der Brennpunkt ein reeller Bildpunkt für einen unendlich weit entfernten, auf der optischen Achse liegenden Objektpunkt. Der Brennpunkt selbst ist reell. Wird der Parabolspiegel als Wölbspiegel verwendet, ist der Brennpunkt virtuell. Beim *elliptischen Hohlspiegel* sind die Brennpunkte reelle Bildpunkte des jeweils anderen Brennpunktes als Objektpunkt. Beim elliptischen Wölbspiegel scheinen Brennpunktstrahlen nach der Reflexion von dem anderen Brennpunkt als virtuellem Bildpunkt auszugehen. Bei der Hohlkugel fallen, wenn man vom Rotationsellipsoid zur Kugel übergeht, beide Brennpunkte im Kugelmittelpunkt zusammen. Der Mittelpunkt bildet sich in sich selbst ab. Beim *Kugelspiegel* wird der Mittelpunkt der Kugelfläche als Brennpunkt aufgegeben. Als Brennpunkt wird der Achsenpunkt definiert, der für einen unendlich weit entfernten Objektpunkt Bildpunkt ist. Wird die ganze Hohlhalbkugel als Spiegelfläche zugelassen, schneiden sich die achsenparallelen Strahlen nach der Reflexion nicht in einem Punkt. Sie haben eine rotationssymmetrische Fläche als Einhüllende, die Brennfläche, deren Meridionalschnitt *Katakaustik* genannt wird. Mit der Beschränkung auf einen achsennahen Raum, das paraxiale Gebiet, entsteht der reelle Bildpunkt als neudefinierter Brennpunkt im Achsenpunkt der Katakaustik in der Mitte zwischen Scheitel und Mittelpunkt der spiegelnden Kugelfläche. Als Wölbspiegel hat der Kugelspiegel eine virtuelle Katakaustik und entsprechend einen virtuellen Brennpunkt hinter dem Scheitel in der Mitte zwischen Scheitel und Mittelpunkt der spiegelnden Kugelfläche. Jede glatt gekrümmte, zur optischen Achse rotationssymmetrische, spiegelnde Fläche hat im Scheitel, ihrem Schnittpunkt mit der optischen Achse, einen bestimmten Krümmungsradius r. Sie kann für den paraxialen Strahlengang durch eine *Schmiegungskugel* mit gleichem Krümmungsradius r_0 ersetzt werden. Von der Kugelform abweichende Spiegelflächen werden demzufolge nur dann eingesetzt, wenn es sich nicht um einen paraxialen Strahlengang handelt, wenn zur Darstellung der brennpunktnahen Bildpunkte durch eine Spiegelung ein homozentrisches Strahlenbündel großer Öffnung zugelassen sein soll und wenn weitere in den Strahlengang eingreifende optische Elemente zur Korrektur der Abbildungsfehler der an sich leichter herzustellenden Kugelfläche nicht zugelassen werden sollen oder können. *Zylinderspiegel* haben eine einfach gekrümmte Spiegelfläche. Als Mantelfläche eines Zylinders kann man sich diese Spiegelfläche durch Parallelverschiebung einer Geraden längs einer in einer Ebene senkrecht zu der Geraden liegenden glatten Leitkurve erzeugt denken. Legt man die Leitkurve in die Zeichenebene, können alle in der Zeichenebene liegenden Komponenten der Lichtstrahlen wie Meridionalstrahlen behandelt werden. Die senkrecht zur Zeichenebene liegenden Komponenten der Lichtstrahlen nehmen an der Spiegelung nicht teil. Bei Zylinderspiegeln gibt es daher kein Problem der windschiefen Strahlen. Ist die Leitkurve ein Kegelschnitt, haben die Zylinderspiegel Brennlinien parallel zu der erzeugenden Mantelgeraden. Ist eine Brennlinie Objektgerade, wird die andere Brennlinie Bildgerade. Ist die Leitkurve ein Halbkreis, hat der zylindrische Halbkreisspiegel eine zylindrische Brennfläche mit einer Katakaustik als Querschnittskurve.

2.5.3.2. Abbildung durch Brechung. *Ebene brechende Flächen* verformen eine von einem Objektpunkt ausgehende Kugelwelle beim Eindringen in das brechende Medium so, daß nur für einen kleinen Öffnungswinkel um das Lot auf die Grenzfläche das divergente Strahlenbündel homozentrisch bleibt. Die rückwärtig verlängerten Strahlen schneiden sich in dem virtuellen Bildpunkt des Objektpunktes.

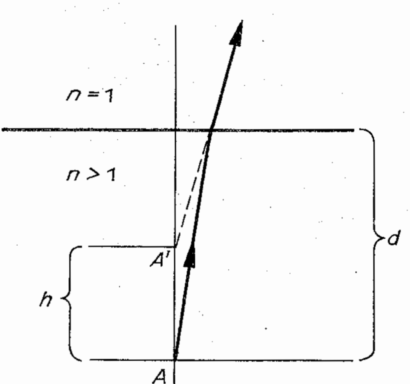


Abb. 2.5.3-3 Bildhebung durch Brechung

Strahlen, die mit dem Lot einen größeren Winkel einschließen, sind nicht mehr ideal-geometrisch abbildend. Tritt eine Kugelwelle vom Objektpunkt A kommend aus dem Medium mit dem Brechungsindex n an dessen ebener Oberfläche aus, hat der Objektpunkt ebenfalls nur ein um

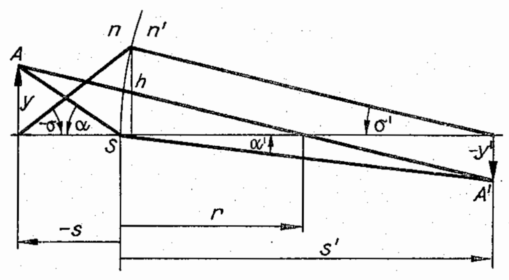


Abb. 2.5.3–4 Abbildung durch brechende Kugelfläche

das Lot eng begrenztes Strahlenbündel, einen virtuellen Bildpunkt A'. Die geometrische Weglänge $AA' = h$ wird als Bildhebung bezeichnet, weil man A' gegenüber A durch das brechende Medium gehoben sieht (Abb. 2.5.3–3). Für kleine Winkel folgt aus dem Brechungsgesetz $d/(d-h) = n$, d. h. eine Bestimmungsgleichung für den Brechungsindex des bildhebenden Mediums.

Brechende Kugelflächen sind die am häufigsten angewendeten abbildenden Elemente. Zum Aufsuchen der Bildpunkte bzw. zur Bildkonstruktion kann man für den paraxialen Strahlengang alle Abbildungsgesetze aus den *Abbeschen Invarianten* ableiten. Es sind dabei allerdings durch Konvention festgelegte Vorzeichenregeln einzuhalten, damit sich das Ergebnis auch allein durch Rechnung einstellt: Eine Strecke ist positiv, wenn sie vom Bezugspunkt aus gesehen rechts liegt. Die Bezugspunkte für die Strecken liegen auf der brechenden bzw. auf der spiegelnden Fläche. Winkel sind positiv, wenn der Lichtstrahl mit der Bezugsgeraden den Winkel von rechts einschließt. Dasselbe gilt für den Radius der Kugelfläche, wenn er mit der optischen Achse als Bezugsgerade einen Winkel einschließt. Für Objektpunkte und für Bildpunkte gelten kartesische Koordinaten, wobei die z-Koordinate in der optischen Achse liegt und nach rechts zeigt. Die y-Koordinate liegt in der Zeichenebene und ist nach oben positiv. Alle Größen und Punkte im Bildraum werden durch Apostrophierung gekennzeichnet. In der optischen Abbildung übernehmen stellvertretend für das gesamte paraxiale homozentrische Strahlenbündel zwei Strahlen die Abbildung von A in A' (Abb. 2.5.3–4). Ein Strahl trifft senkrecht auf die Kugelfläche mit dem Radius r und geht ungebrochen bis zum Bildpunkt A'. Der andere Strahl wird im Scheitel S durch die Kugelfläche mit ihrer senkrecht zur optischen Achse orientierten Tangentialebene zur optischen Achse hin gebrochen. Tangens, Sinus und Winkel werden im paraxialen Strahlengang gleich gesetzt. Nach dem Brechungsgesetz gilt: $n'/n = \sin\alpha/\sin\alpha' = \tan\alpha/\tan\alpha' = [y/(-s)]/(-y'/s')$. Aus ihm leitet sich der *Abbildungsmaßstab* $\beta' = y'/y = ns'/(n's)$ bzw. die Vergrößerung $\Gamma = s/s' = \tan\sigma'/\tan\sigma \approx \sigma'/\sigma = ny/(n'y')$ ab. Γ wird auch als *Winkel-* oder *Angularvergrößerung* bezeichnet. In der Form $ny\sigma = n'y'\sigma'$ wird diese Aussage auch als *Helmholtz-Gleichung* bezeichnet, d. h., das Produkt aus Brechzahlverhältnis, Abbildungsmaßstab und Winkelvergrößerung ist gleich Eins. Nach dem Strahlensatz gilt $y/(-s+r) = -y'/(s'-r)$. Mit dem Brechungsgesetz folgt die *Abbesche Invariante* $n(1/r - 1/s) = n'(1/r - 1/s')$. Aus der Abbeschen Invariante folgt mit $s' = \infty$ bzw. $s = -\infty$ die objekt- bzw. die bildseitige *Brennweite* zu $f = -nr/(n'-n)$ und $f' = n'r/(n'-n)$. Die Beträge der Brennweiten verhalten sich zueinander wie die Brechungsindizes. Führt man die Brennweiten in die Abbesche Invariante als allgemeine Abbildungsgleichung ein, vereinfacht sich die Abbildungsgleichung zu $f'/s' + f/s = 1$. Legt man die Brennpunkte als die Koordinatenursprünge im Objekt- und Bildraum fest, d. h. setzt man $s = z + f$ bzw. $s' = z' + f'$, dann kann man die Abbildungsgleichung auch $zz' = ff'$ schreiben. Sie wird die *Newtonsche Abbildungsgleichung* genannt.

Bei der ideal-geometrischen Abbildung werden die in einer Ebene senkrecht zur optischen Achse liegenden achsennahen Objektpunkte in der sogenannten *Gaußschen Bildebene*, die senkrecht zur optischen Achse durch den paraxialen Bildpunkt geht, abgebildet. Objektebene und Bildebene sind zueinander konjugiert. Eine zweite Objektebene hat eine entsprechende zu ihr konjugierte Bildebene. Das Verhältnis des Abstandes der beiden Bildebenen zum Abstand der beiden Objektebenen gibt Auskunft über die Abbildung der senkrecht zur Objektebene vorliegenden Ausdehnung des Objektes und wird *Axialvergrößerung* genannt. Sie ist dem Quadrat des Abbildungsmaßstabes proportional. Der Proportionalitätsfaktor ist das Brechungsverhältnis. Eine objektähnliche Abbildung erfahren nur die in einer Ebene senkrecht zur optischen Achse liegenden Strukturen des Objektes.

Setzt man in der Abbeschen Invarianten $n' = -n$, spiegelt man den Bildraum in den Objektraum und die Abbesche Invariante wird zur Abbildungsgleichung für den rotationssymmetrischen Spiegel bei paraxialem Strahlengang, $1/r - 1/s = -1/r + 1/s'$. Mit $s = -\infty$ ergibt sich für den Hohlspiegel ($r < 0$) eine Brennweite $f' = s' = -r/2$. Wird die Objektschnittweite, der Abstand vom Scheitel bis zum Schnittpunkt des Lichtstrahles mit der optischen Achse s, positiv, wird der Kugelspiegel zum Wölbspiegel. Mit $s = \infty$ ist die Brennweite $f' = s' = -r/2$. Der Brennpunkt ist jedoch virtuell. Reeller und virtueller Brennpunkt fallen zusammen. Das Abbildungsgesetz nimmt die einfache Form $1/s + 1/s' = 1/f = 1/f'$ mit $f = f' = r/2$ an. Für den ebenen Spiegel ist der Krümmungsradius $r = \infty$. Aus der Abbildungsgleichung folgt $s' = -s$, d. h., der Bildpunkt hat von der Spiegelebene dieselbe Entfernung wie der Objektpunkt, liegt aber auf der anderen Seite der Spiegelebene.

Zur Darstellung eines Bildpunktes kann man sich auch sogenannter ausgezeichneter Strahlen bedienen, deren Verlauf durch Kardinalelemente festgelegt ist (Abb. 2.5.3–5). Für ein zentriertes System von spiegelnden und brechenden Flächen sind dies drei Paare ausgezeichneter Punkte auf der optischen Achse des Systems, die auch Kardinalpunkte genannt werden: *Brennpunkte*, *Hauptpunkte* und *Knotenpunkte*. Sie sind wie folgt definiert: Ein der optischen

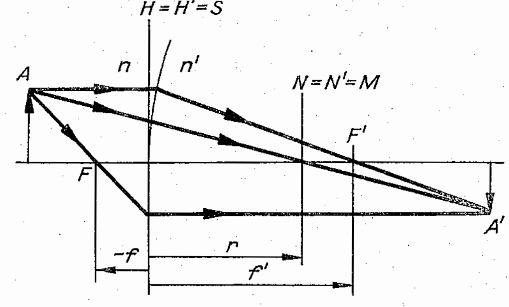


Abb. 2.5.3–5 Kardinalelemente und ausgezeichnete Strahlen bei Abbildung durch brechende Kugelfläche

Achse parallel einfallender Strahl schneidet die optische Achse nach dem Durchgang durch das System im bildseitigen Brennpunkt F'. Die Verlängerung des achsenparallelen Strahles schneidet die rückwärtige Verlängerung des das System verlassenden Strahles in der durch den Hauptpunkt gehenden achsensenkrechten Ebene, der bildseitigen Hauptebene H'. Ein von einem außeraxialen Objektpunkt ausgehender Strahl, der nach Durchgang durch das System dieses in derselben Richtung, also ungebrochen, gegebenenfalls parallel versetzt, verläßt, schneidet die optische Achse im objektseitigen Knotenpunkt N. Er scheint nach Verlassen des Systems vom bildseitigen Knotenpunkt N' auszugehen. Ein das System parallel zur optischen Achse verlassender Strahl geht von einem außeraxial gelegenen Objektpunkt aus und schneidet vor Eintritt in das System die optische Achse im objektseitigen Brennpunkt F. Die Verlängerung des Brennpunktstrahles und die rückwärtige Verlängerung des das System parallel zur optischen Achse verlassenden Strahles schneiden sich in der durch den objektseitigen Hauptpunkt gehenden achsensenkrechten Ebene, der objektseitigen Hauptebene H. Hauptpunkte und -ebenen werden i. allg. mit demselben Buchstaben bezeichnet. Mit der Kenntnis der Kardinalpunkte kann man besonders vorteilhaft bei zusammengesetzten, zentrierten, optisch abbildenden Systemen stellvertretend für den wirklichen Strahlengang einen über alles zusammenfassenden Strahlengang von zwei ausgezeichneten Strahlen mit einem Hauptebenenpaar, den Knotenpunkten, und den beiden Brennpunkten des Systems zur Bildkonstruktion bzw. -berechnung benutzen. Bei gleichem Medium fallen die Knotenpunkte vor und hinter dem System mit den entsprechenden Hauptpunkten zusammen.

2.5.3.3. Linsen. Unter einer Linse versteht man ein von zwei brechenden Flächen begrenztes optisches System mit abbildenden Eigenschaften. Die brechenden Flächen sind i. allg. Kugelflächen. Torische, zylindrische und rotationssymmetrische asphärische Flächen werden nur in Ausnahmefällen verwendet. Linsen bestehen meist aus Glas, auf jeden Fall aber aus einem optisch isotropen Medium. Linsen unterscheiden sich außer durch das Material, aus dem sie hergestellt sind, noch durch ihre Form, die Radien der brechenden Flächen, ihre Mittendicke und durch ihren Durchmesser. Die allgemeinen Formen sind plankonvex, plankonkav, bikonvex, bikonkav und konvexkonkav. Konvexkonkavlinsen heißen auch Meniskuslinsen. Bikonvex-, Plankonvex- und sammelnde Meniskuslinsen sind durch die kreisförmige Schnittkurve ihrer brechenden Flächen im Außendurchmesser begrenzt, alle anderen Linsen durch eine Facettierung. Trotzdem erhalten auch die erstgenannten zur besseren Handhabung eine Außenfacettierung, die den für die Abbildung zur Verfügung stehenden Durchmesser verkleinert und wie bei den anderen Linsen eine Formachse festlegt. Formachse und optische Achse müssen bei zentrierten Systemen zusammenfallen.

Zur Ableitung der *Abbildungsgesetze* einer Linse hat man die Abbildungsgesetze für brechende Flächen zweimal nacheinander auf die brechenden Flächen der Linse anzuwenden. Man übernimmt mit den Abbildungsgesetzen auch die Beschränkung des Strahlenganges auf ein paraxiales Gebiet. Die *Brennweiten* ergeben sich als Abstand zwischen Brennpunkt und zugehörigem Hauptpunkt bild- bzw. objektseitig zu $f' = -f'_1 f'_2/t$ bzw. $f = f_1 f_2/t$. Die Indizes 1 und 2 beziehen sich auf die objektseitige bzw. bildseitige brechende Fläche, t ist die *optische Tubuslänge* oder das *optische Intervall* als Abstand zwischen dem bildseitigen Brennpunkt der ersten Fläche zum objektseitigen Brennpunkt der zweiten Fläche. Der Abstand des objektseitigen Brennpunktes der Linse vom objektseitigen Brennpunkt der ersten Fläche ist $z_1 = f_1 f'_1/t$. Entsprechend gilt für den bildseitigen Brennpunkt der Linse $z'_2 = -f_2 f'_2/t$. Befindet sich vor und hinter der Linse Luft ($n = 1$), ergeben sich bild- und objektseitige Brennweite zu $f' = -f = n r_1 r_2/(n-1)R$ mit $R = n(r_2 - r_1) + (n-1)d$. Hierin bedeuten r_1 und r_2 die Radien der brechenden Kugelflächen der Linse, n den Brechungsindex des Linsenmateriales und d die Mittendicke der Linse. Die beiden Brennweiten sind dem Betrage nach gleich. Das Minuszeichen weist nur aus, daß die beiden Brennpunkte auf verschiedenen Seiten der Linse liegen. Entsprechend gilt für die Lage der Hauptpunkte mit e als Abstand vom zugehörigen Linsenscheitel $e = -r_1 d/R$ und $e' = -r_2 d/R$. Da man für eine Planfläche $r = \infty$ setzen kann, folgt für die plankonvexe wie für die plankonkave Linse, daß immer die eine Hauptebene den Scheitel der Kugelfläche tangiert und die andere um den Betrag d/n von der Planfläche aus nach innen gerückt liegt. Werden die Abstände von Objekt- und Bildebene in Hauptpunktkoordinaten a und a' dargestellt, lautet die Abbildungsgleichung unter Beachtung der Vorzeichenvereinbarung $1/a' - 1/a = 1/f'$, in Worten: Die Summe aus den Kehrwerten von Gegenstands- und Bildweite ist gleich dem Kehrwert der Brennweite bzw. gleich der *Brechkraft der Linse*. Die Brechkraft wird in Dioptrien (dpt) angegeben, die Brennweite ist dabei in Meter einzusetzen. Sie ist für Sammellinsen > 0, für Zerstreuungslinsen < 0. Eine Linse mit der Brennweite $f' = -50$ cm hat eine Brechkraft $1/(-0{,}5\,\text{m}) = -2\,\text{m}^{-1} = -2$ dpt. Werden die Abstände von Objekt- und Bildebene auf die Brennpunkte bezogen, nehmen die Abbildungsgleichungen die folgende einfache Form $zz' = -f'^2$ und $\beta' = -f/z = -z'/f'$ an. In dieser Form geschrieben werden sie auch *Gaußsche Abbildungsgesetze* genannt. Einen Sonderfall stellt die telezentrische symmetrische Bikonvexlinse dar. Bei ihr fallen der bildseitige Brennpunkt der ersten Fläche und der objektseitige Brennpunkt der zweiten Fläche zusammen. Ein achsenparalleler einfallender Strahl verläßt die Linse achsenparallel. $\beta' = -1$ unabhängig von der objektseitigen Scheitelweite. Der Abstand Objektebene – Bildebene ist $2r(n+1)/(n-1) = \text{konst.} = 2(f'_1 + f'_2)$. Nur im Scheitelweitenintervall $(-2f_1; 0)$ liegende Objektpunkte werden reell abgebildet.

Bei einer *Kugellinse* ist der Mittelpunkt der Kugel auch Knotenpunkt, d. h. die beiden Knotenpunkte N und N' fallen zusammen. Damit fallen auch die Hauptebenen zusammen und gehen durch den Mittelpunkt der Kugel. Brenn-, Gegenstands- und Bildweite sind vom Kugelmittelpunkt aus zu zählen. Die Brennweite ergibt sich zu $f = -nr/2(n-1)$. Für $n = 2$ folgt daraus, daß die beiden Brennpunkte auf der Oberfläche der Kugel liegen. Als dünnwandige, wassergefüllte Glaskugel (Schusterkugel) ist wegen $n = 1{,}33$ die Brennweite gleich dem doppelten Radius und für eine Vollglaskugel mit $n = 1{,}5$ gleich dem 1 1/2fachen Radius. Bei einer dünnen Linse fallen Haupt-, Knoten- und Scheitelpunkte zusammen. Als *Hoeghscher Meniskus* wird eine Linse mit zwei gleichen Krümmungsradien bezeichnet. Seine Brechkraft ist der Mittendicke d proportional, $1/f' = (n-1)^2 d/nr^2$.

Zylinderlinsen haben eine einfach gekrümmte, brechende Fläche. Diese ist i. allg. die Mantelfläche eines geraden Kreiszylinders. Die zweite, eine Zylinderlinse begrenzende, brechende Fläche ist i. allg. eben. Der Schnitt senkrecht zur Zylinderachse ist der wirksame Schnitt. Für Strahlen in die-

ser Ebene gelten alle für Meridionalstrahlen an zentrierten brechenden Systemen aufgestellten Gesetze. Entsprechend haben Zylinderlinsen eine Brechkraft und eine Brennweite für diese Strahlen. Die gegen einen wirksamen Schnitt geneigt verlaufenden Strahlen sind windschief, sofern sie nicht in der Ebene verlaufen, die die zu den brechenden Flächen gehörenden Achsen bzw. deren Scheitelgeraden enthält oder, wenn die zweite brechende Fläche eine ebene Fläche ist, senkrecht dazu orientiert ist und die Achse bzw. die Scheitelgerade der anderen brechenden Fläche enthält. Für paraxiale Strahlen in dieser Ebene kann man den Abbildungsmaßstab $\beta' = 1$ setzen. Zylinderlinsen werden meist nur in Verbindung mit einem zentrierten abbildenden System benutzt. Sie bewirken – während das zentrierte optische System die Punktabbildung übernimmt – in der Richtung senkrecht zu ihrer Zylinderachse eine Stauchung oder Streckung des Bildes. Mit zwei Zylinderlinsen, deren Brennweite groß gegen ihre Mittendicke ist, läßt sich bei rechtwinklig gekreuzter Aufstellung eine punktförmige Abbildung erreichen. Bei gleicher Brennweite können zwei Zylinderlinsen in gekreuzter Stellung durch eine ersetzt werden, deren brechende Flächen senkrecht gekreuzt den Linsenkörper begrenzen. Werden zwei Zylinderlinsen mit unterschiedlicher Brennweite so in gekreuzter Stellung mit unterschiedlicher Gegenstandsweite aufgestellt, daß die Bildfokussierung durch die wirksamen Schnitte in der gleichen Bildebene erfolgt, kommt es zur *anamorphotischen Abbildung*, d. h. zu einer Abbildung mit zwei rechtwinklig zueinander unterschiedlichen Abbildungsmaßstäben. Diese Möglichkeit wird bei der Breitwandkinematografie genutzt. Die Filmaufnahme auf Film mit Normalformat erfolgt mit Stauchung in der Horizontalen, die Projektion auf die Filmleinwand mit entzerrender Streckung in der Horizontalen. Die Zusatzsysteme für solche Aufnahmen bzw. Projektion heißen Anamorphot. Eine *torische Fläche* ist als brechende Fläche einer Linse die Spur einer Meridionalkurve nach Rotation um eine Achse, die im Meridionalschnitt senkrecht zur optischen Achse liegt. Ist die Meridionalkurve ein Kreisbogen, hat die torische Fläche im Meridionalschnitt entweder einen größeren (tonnenförmige torische Fläche) oder einen kleineren (wurstförmige torische Fläche) Krümmungsradius als im senkrecht dazu liegenden Sagittalschnitt. Die zweite brechende Fläche einer Linse mit torischer Fläche ist i. allg. eine Kugelfläche. Linsen dieser Art dienen hauptsächlich als Brillengläser zur Korrektur eines Astigmatismus (vgl. 2.5.3.4.) der Augenlinse.

2.5.3.4. Abbildungsfehler. Abbildungsfehler als Abweichung von der ideal-geometrischen Abbildung haben als Ursache die Dispersion der optischen Werkstoffe und die Zulassung von Strahlen außerhalb des paraxialen Gebietes. Sie werden deshalb unterteilt in *Farbfehler* und *geometrische Fehler*. Bei nur spiegelnden Systemen treten Farbfehler nicht auf. Die Abbildungsfehler werden in der geometrischen Optik quantitativ als Strahlabweichungen oder *Aberrationen* erfaßt. Zur Darstellung der Querabweichungen oder -fehler wird der Durchstoßpunkt eines Bezugsstrahles mit den Koordinaten x', y' durch eine achssenkrechte Bezugsebene vorgegeben. Die *sagittale Querabweichung* ist dann $\Delta x' = \hat{x}' - x'$, die *meridionale Querabweichung* $\Delta y' = \hat{y}' - y'$ mit $\hat{x}'$ und $\hat{y}'$ als Koordinaten des Strahldurchstoßpunktes durch die Bezugsebene. Zur Darstellung der *Längsabweichungen* wird ein Bezugsstrahl mit seinem Durchstoßpunkt durch eine achssenkrechte Bezugsebene vorgegeben. Die Längsabweichung ist dann entweder der Senkrechtabstand des Schnittpunktes von Strahl und Bezugsstrahl von der Bezugsebene oder der Abstand auf dem Bezugsstrahl vom Durchstoßpunkt durch die Bezugsebene. Die grafische Darstellung der Abbildungsfehler erfolgt so, daß die Abszisse Längsfehlerkoordinate, die Ordinate Querfehlerkoordinate ist. Die verursachenden Parameter werden mit der jeweils anderen Koordinate vorgegeben. Bezugsstrahl ist ein Hauptstrahl oder die optische Achse. Ist die optische Achse Bezugsstrahl, wird der Längsfehler als Schnittweitendifferenz angegeben, $\Delta s' = \hat{s}' - s'$ (Abb. 2.5.3–6). Bezugsebene ist die Gaußsche Bildebene.

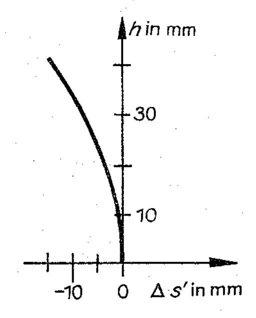


Abb. 2.5.3–6 Sphärische Aberration; Längsfehler als Schnittweitendifferenz in Abhängigkeit von der Einfallshöhe (vgl. Abb. 2.5.3–4)

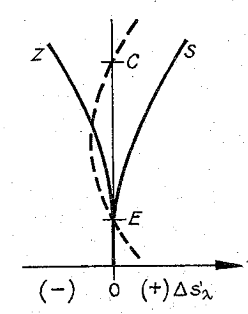


Abb. 2.5.3–7 Chromatische Aberration; —— primäres Farbfehlerspektrum einer Sammel- bzw. Zerstreuungslinse, – – – sekundäres Farbfehlerspektrum eines Dichromaten („Achromat" bei gleichzeitiger Korrektur der sphärischen Aberration)

Der *Farblängsfehler* oder die *chromatische Aberration* tritt bereits im paraxialen Gebiet auf. Bei Sammellinsen wird mit dem Übergang von violettem zu rotem Licht infolge der normalen Dispersion des Linsenmateriales die Schnittweite der Gaußschen Bildebene größer. Das Farbfehlerspektrum ist ein primäres. Bei Zerstreuungslinsen ist die Farbfolge umgekehrt (Abb. 2.5.3–7). Bei der Abbildung von außeraxialen Objektpunkten kommt es zusätzlich zu einem mit der Wellenlänge des Lichtes veränderlichen Abbildungsmaßstab $\beta'(\lambda)$. Die chromatische Aberration läßt sich mit zwei Linsen durch Einhalten der *Achromasiebedingung* $\nu_1 f_1' + \nu_2 f_2' = 0$ für zwei Farben beseitigen. Die ν_i sind die Abbeschen Zahlen des Linsenmateriales, die f_i' die Brennweiten. Eine Linse muß eine Zerstreuungslinse, die andere eine Sammellinse sein. Es ist üblich, die Achromasie für die beiden Farben der Fraunhoferschen Linien C und F herzustellen (vgl. Abb. 2.5.3–7). Die Restfehler werden als das sekun-

däre Spektrum bezeichnet. Bei der grafischen Darstellung wird die Schnittweite für die Einstellebene gleich Null gesetzt und der Farblängsfehler $\Delta s'_\lambda$ auf diese Einstellebene bezogen. Er läßt sich bei Verwendung von drei Linsen weiter einschränken, indem man die Achromasiebedingung zweimal anwendet und eine Übereinstimmung der Bildebene für drei Farben herstellt (Abb. 2.5.3–8). Die verbleibenden Restfehler werden das *tertiäre Spektrum*, das so korrigierte Linsensystem ein *Apochromat* genannt. Die geometrischen Fehler der Linsen sind für monochromatisches Licht erklärt. Es ist deshalb stets mit einer Wellenlängenabhängigkeit dieser Fehler zu rechnen. Sie wird als chromatische Variation der geometrischen Fehler bezeichnet. Die chromatische Variation des Öffnungsfehlers ist der sogenannte Gauß-Fehler. Der *Öffnungsfehler* oder die *sphärische Aberration* ist als Längsfehler die Abweichung der bildseitigen Schnittweite für Strahlen mit der Einfallshöhe h von der für paraxiale Strahlen $\Delta s'_h = \hat{s}' - s'$ (vgl. Abb. 2.5.3–6). Achsenparallele Strahlen schneiden sich als achsennahe Strahlen im *paraxialen Brennpunkt*, als Randstrahlen mit der größtmöglichen Einfallshöhe h im *marginalen Brennpunkt*.

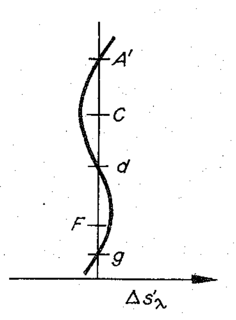


Abb. 2.5.3–8 Chromatische Aberration; tertiäres Spektrum eines Trichromaten („Apochromat" bei gleichzeitiger Korrektur des Gauß-Fehlers)

Die bildseitigen Strahlen tangieren alle eine rotationssymmetrische Fläche, die in den paraxialen Brennpunkt ausläuft und als Einhüllende die Brennfläche oder die Kaustik genannt wird. Sie ist deshalb Brennfläche, weil sie durch die Schnittkreise der benachbarten bildseitigen Strahlenkegel dargestellt wird. Es schneiden sich aber auch die zu einem Strahlenkegel gehörenden bildseitigen Strahlen in einem Achsenpunkt, so daß gleichzeitig mit der Brennfläche eine Brennlinie auf der optischen Achse zwischen marginalem und paraxialem Brennpunkt entsteht. Das Strahlenbündel erfährt in diesem Bereich nur eine Einschnürung. Legt man die Einstellebene zum Auffangen des Bildes an die Stelle mit der engsten Einschnürung, ist die Intensität des Lichtes über den Querschnitt des zu einem *Zerstreuungskreis* gewordenen Bildpunktes nahezu gleichverteilt ohne Nebenlicht außerhalb. Bei der Abbildung grober Strukturen wird dadurch der Kontrast größer als bei einer Einstellebene, bei der im Zentrum des Streuungskreises die größte Helligkeit wellenoptisch zu erwarten ist. Mit einer solchen Einstellebene läßt sich jedoch ein größeres Auflösungsvermögen erzielen. Ihr Abstand vom paraxialen Brennpunkt ist $b' = 0{,}5\Delta s'_{max}$. Sie liegt also in der Mitte zwischen marginalem und paraxialem Brennpunkt, während die Stelle engster Einschnürung näher zum marginalen Brennpunkt liegt. Der Öffnungsfehler einer Linse nimmt mit größer werdendem Brechungsindex des Linsenmaterials ab. Er ist immer größer als der Öffnungsfehler eines vergleichbaren Kugelspiegels. Nach Korrektur des Öffnungsfehlers für Strahlen mit großer Einfallshöhe h verbleiben für Strahlen mit geringerer Einfallshöhe zentrierte Bereiche oder Zonen, in denen ein Restfehler der sphärischen Aberration als *Zonenfehler* vorhanden ist. Der *Zweischalenfehler* oder *Astigmatismus schiefer Bündel* ist die unterschiedliche Längsabweichung der Bildpunkte von außeraxialen Objektpunkten in der Meridional- und Sagittalebene. Er wird als Längsfehler $b'_{m,s}$ in Abhängigkeit vom Schnittwinkel des Hauptstrahles mit der optischen Achse dargestellt. Die Bildpunkte der Objektpunkte einer Objektebene liegen somit auf zwei getrennten rotationssymmetrisch zur optischen Achse liegenden Schalen, die nur den Gaußschen Bildpunkt auf der optischen Achse gemeinsam haben. Ein kreisförmiger Querschnitt eines von einem außeraxialen Objektpunkt ausgehenden Strahlenbündels ändert sich nach Durchgang durch die Linse kontinuierlich zu einem elliptischen, um dann in einen strichförmigen überzugehen. An dieser Stelle entsteht in der Meridionalebene ein „Bildpunkt", der aber senkrecht dazu in der Sagittalebene als Bildlinie ausgedehnt ist. Mit allen solchen Bildpunkten wird die meridionale Bildschale dargestellt. Im weiteren Verlauf des Strahlenganges geht der Querschnitt des bildseitigen Strahlenbündels vom strichförmigen in einen kreisförmigen über. Dort hat das Strahlenbündel seine engste Einschnürung. Danach wird der Querschnitt des Strahlenbündels wieder strichförmig. Es kommt zu einem „Bildpunkt" in der Sagittalebene, der senkrecht dazu in der Meridionalebene als Bildlinie ausgedehnt ist. Mit allen solchen Bildpunkten wird die sagittale Bildschale aufgebaut. Die beiden außeraxialen Längsabweichungen werden auch als *meridionale* und *sagittale Bildfeldkrümmung* bezeichnet, ihre Differenz als *astigmatische Differenz*. Durch die räumliche Ausdehnung der Bildpunkte zu Bildlinien bewirkt der Astigmatismus eine charakteristische Unschärfe mit zunehmender Entfernung von der Bildmitte derart, daß entweder nur in radialer Richtung, d. h. in der Meridionalebene, orientierte Objektstrukturen scharf abgebildet werden können oder nur in tangentialer Richtung, d. h. in der Sagittalebene, orientierte (Abb. 2.5.3–9). Nur für die Bildmitte gelingt die Einstellung einer von der Orientierung der Objektstrukturen unabhängigen Bildschärfe. Die Beseitigung des Zweischalenfehlers durch Beseitigen der astigmatischen Differenz führt zur *Bildfeldwölbung*. Alle Objektpunkte eines ebenen Objektes werden nach Beseitigung der astigmatischen Differenz auf einer zur optischen Achse rotationssymmetrischen, bei Sammellinsen zum Objekt hin konkaven Schale abgebildet, die *Petzval-Fläche* genannt wird. Für eine dünne beidseitig von Luft begrenzte Linse ist der Krümmungsradius dieser Fläche $r_P = -nf'$. Für mehrere gemeinsam wirkende dünne Linsen ist der Kehrwert des Krümmungsradius der Petzval-Fläche des Linsensystems gleich der Summe der Kehrwerte der Krümmungsradien r_P der Einzellinsen, $1/r_P = -\sum 1/n_i f'_i$. Für eine Folge von k brechenden Flächen ist $1/r_P = -n'_k \sum -(1/n'_i - 1/n_i)/r_i$. Hierin sind die n'_i jeweils der Brechungsindex des an die i-te Fläche angrenzenden Mediums. Bei einer spiegelnden Fläche ist $n' = -n$ zu setzen. Die die Krümmung der Petzval-Fläche bestimmende Summe wird Petzval-Summe genannt. Je kleiner sie ist, umso besser ist die Petzval-Fläche als korrigierte Bildfläche eingeebnet. Mikroskopobjektive,

bei denen die Korrektur der Abbildungsfehler die Bildebnung einschließt, werden durch die Vorsilbe Plan- gekennzeichnet, z. B. Planapochromat usw.
Die *Koma* ist das geschweifte Bild eines außeraxialen Objektpunktes. Es ist spiegelsymmetrisch zum Meridionalschnitt. Die Koma hat ihre Ursache darin, daß das vom Objektpunkt ausgehende rotationssymmetrische Strahlenbündel die einzelnen Zonen des nicht oder nicht einheitlich sphärisch korrigierten Systems im Meridionalschnitt unsymmetrisch verteilt durchsetzt. Dadurch schneiden sich alle in der Meridionalebene zur gleichen Öffnungszone gehörenden Strahlen nach der Brechung seitlich vom Hauptstrahl. Die Zerstreuungskegel bilden in der Meridionalebene eine Brennlinie, deren Durchstoßpunkt durch die Bildebene als der Bildpunkt des Objektpunktes angesehen wird, weil er der hellste Teil der Koma ist und dem Gaußschen Bildpunkt am nächsten kommt. Der Zerstreuungskreis weist dann als sogenannter Komaschweif in der Bild-

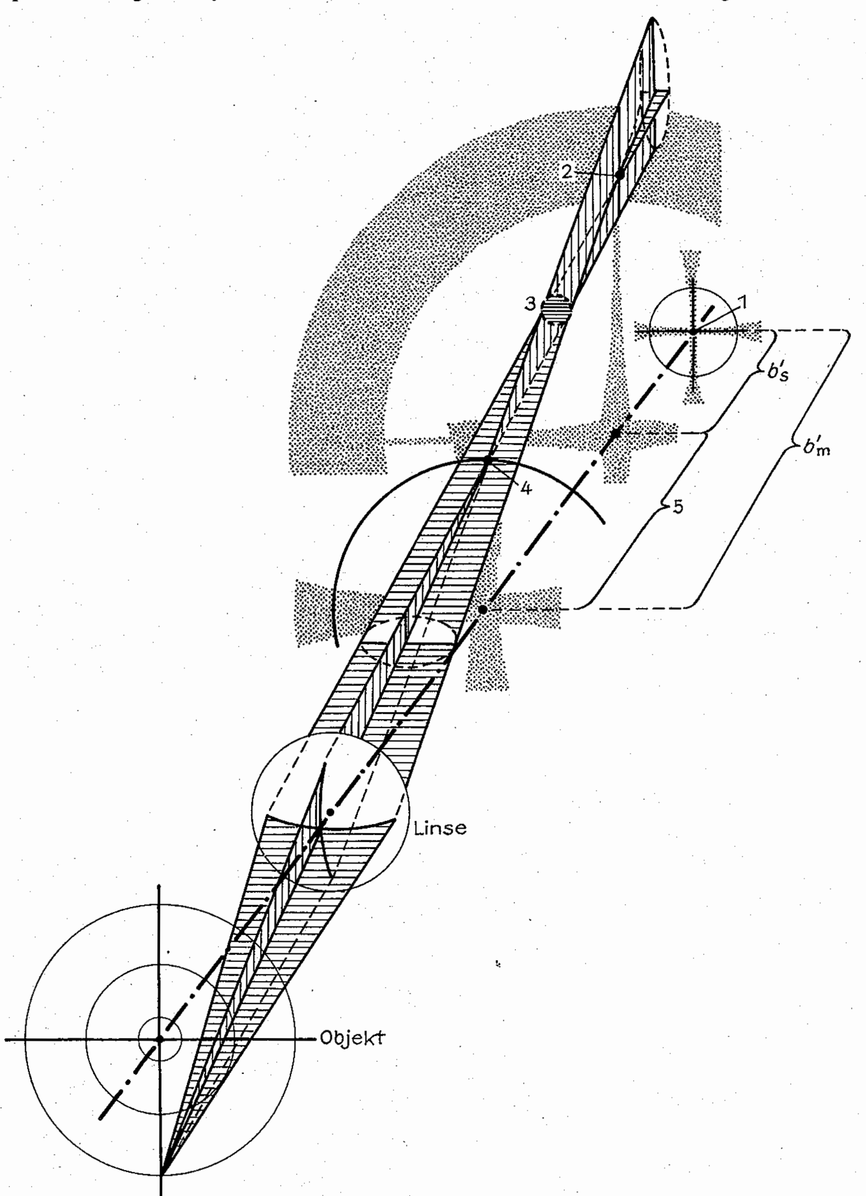


Abb. 2.5.3–9 Zweischalenfehler oder Astigmatismus schiefer Bündel; *1* Achsenpunkt der Gaußschen Bildebene, *2* Bildpunkt der sagittalen Bildschale, *3* kleinster Zerstreuungskreis, *4* Bildpunkt der meridionalen Bildschale, *5* astigmatische Differenz, b'_m bzw. b'_s meridionale bzw. sagittale Bildfeldkrümmung

ebene zur optischen Achse hin oder von ihr weg. Entsprechend nennt man die Koma Innen- bzw. Außenkoma. Die *Verzeichnung* ist eine radiale Querabweichung der Bildpunkte von den mit dem paraxialen Abbildungsmaßstab vorgegebenen Bildpunktlagen. Quantitativ wird die Verzeichnung als das prozentuale Verhältnis der Abweichung $\Delta y'$ von der paraxial gegebenen Bildgröße y' angegeben. Ist die Verzeichnung > 0, nimmt ein abzubildendes Quadrat eine Kissenform an, ist sie < 0, eine Tonnenform. Man spricht deshalb von Kissen- oder Tonnenverzeichnung. Soll eine Abbildung wie in der Fotogrammetrie streng zentralperspektivisch erfolgen, ist entweder Verzeichnungsfreiheit des abbildenden Objektives zu fordern oder die Verzeichnung ist bekannt und kann berücksichtigt oder durch entsprechende Korrektionsplatten bei der Bildauswertung wieder rückgängig gemacht werden. Die Abbildung durch einfache dünne Linsen oder durch symmetrische Systeme mit zentralgelegener Öffnungsblende führt nicht zu einer Verzeichnung. Das Auftreten einer Verzeichnung bei Verwendung einer Vorder- oder Hinterblende oder eines asymmetrisch zusammengestellten bzw. aufgebauten Systems läßt sich dadurch erklären, daß, wenn vorhanden, eine Korrektur der sphärischen Aberration strenggenommen nur für eine Objekt- bzw. Bildweite möglich ist und nicht gleichzeitig auch für die Abbildung einer Öffnungsblende als Eintritts- bzw. Austrittspupille in gleichem Ausmaß wirksam sein kann. Bei einer Sammellinse wird deshalb mit zunehmendem Schnittwinkel des durch die Mitte der Öffnungsblende gehenden Hauptstrahles auch das Bild der Öffnungsblende sich von ihrem paraxialen Bild und von der Linse weiter entfernen. Ein System ist nur dann verzeichnungsfrei, wenn die Austrittspupille für alle Bildpunkte und die Eintrittspupille für alle Objektpunkte die gleiche Entfernung von der Bild- bzw. Objektebene hat. Diese Voraussetzung schließt auch die *Tangensbedingung* ein, die bei Verzeichnungsfreiheit erfüllt ist. Sie lautet für alle das Objekt abbildenden Strahlen $\tan \sigma_i' / \tan \sigma_i =$ konst.

2.5.3.5. Zusammengesetzte optische Systeme. Ein optisches System ist eine Folge von abbildenden u. a. optischen Funktionselementen mit einer optischen Gesamtwirkung. Diese optische Gesamtwirkung dient hauptsächlich der optischen Abbildung. Sie kann auch der Beleuchtung, der physikalischen Untersuchung, gegebenenfalls in Verbin-

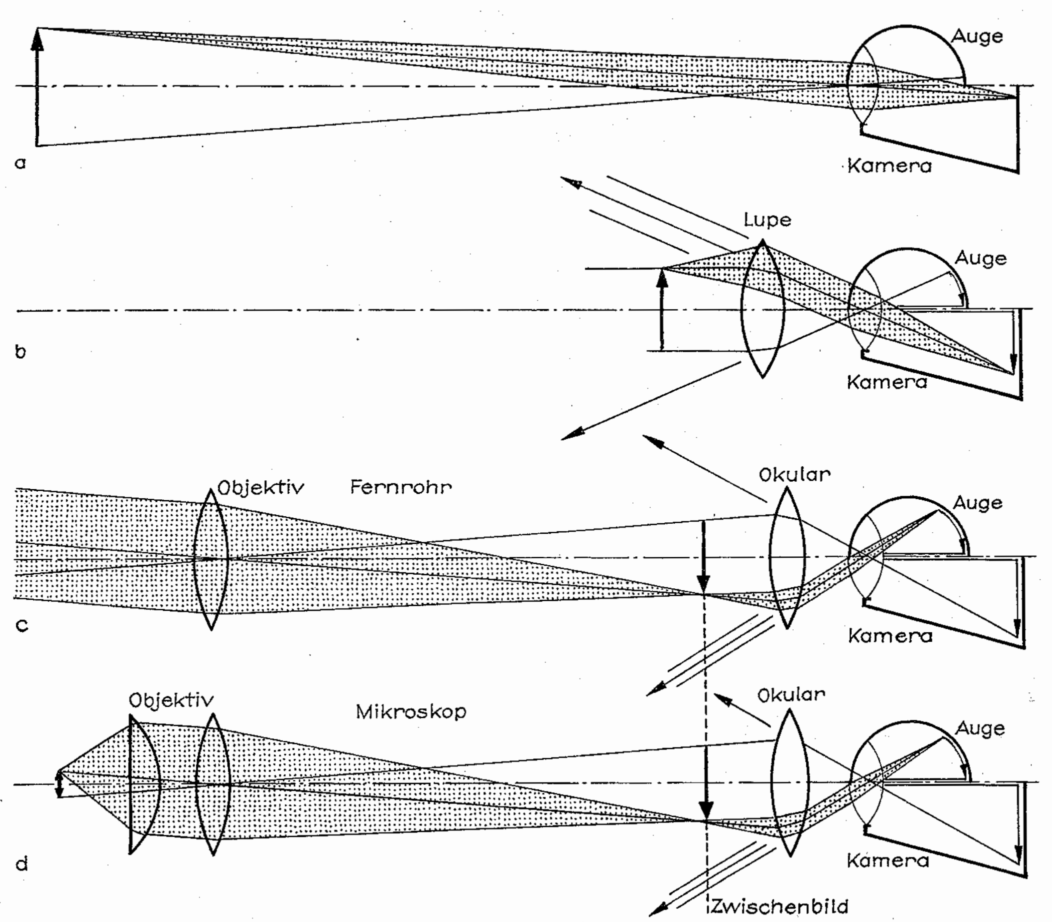


Abb. 2.5.3–10 Überblick über die verschiedenen Arten der optischen Geräte (nach MICHEL), (a) Auge und fotografische Kamera, (b) Lupe und Vorsatzlinse, (c) Fernrohr und Fernrohrkamera, (d) Mikroskop und mikrofotografische Kamera

dung mit einer Abbildung, oder der Fernleitung von Licht als Informationsträger dienen. Bei optischen Systemen für Zwecke der Abbildung wird die Zusammensetzung des Systems von den an die Abbildung zu stellenden Forderungen bestimmt. Diese beziehen sich im wesentlichen auf die Öffnung des Systems, auf die Größe des Gesichtsfeldes, auf die Vergrößerung und auf das Auflösungsvermögen. Im Zusammenhang damit entstehen Forderungen in bezug auf zu beseitigende Abbildungsfehler.
Bei den verdeutlichenden Geräten bilden das Auge und das optische Gerät eine funktionelle Einheit. Bei den reproduzierenden Geräten tritt an die Stelle des Auges eine Kamera, die auch schon für sich betrachtet ein reproduzierendes Gerät ist. Einen Überblick über die verschiedenen Arten der optischen Geräte zeigt Abb. 2.5.3–10.
Zu den verschiedenen zusammengesetzten optischen Systemen vgl. 7.5.

2.6. Wärmelehre

Die physikalischen Systeme, mit denen man im täglichen Leben zu tun hat, bestehen aus sehr vielen Teilchen, vielen Atomen oder Molekülen. Für diese Systeme lassen sich makroskopische Gesetzmäßigkeiten formulieren, ohne auf die Bewegung der einzelnen Bestandteile eingehen zu müssen. In diesen Gesetzen tauchen nur wenige makroskopische Parameter, wie Temperatur, Volumen und Druck, auf. Diese Gesetzmäßigkeiten und die daraus folgenden Aussagen nennt man Wärmelehre, in ihrer stärker theoretischen Durchdringung *Thermodynamik*. Die Temperatur ist der für die Wärmelehre charakteristische Parameter. Die Zusammenhänge zwischen den Gesetzmäßigkeiten der Wärmelehre und den mikroskopischen Bewegungsgleichungen der einzelnen Teilchen werden in der *Statistik* (vgl. 4.8.) untersucht. Die Aussagen der Wärmelehre sind außerordentlich weitreichend. Einzige Voraussetzung ist, daß das betrachtete Gesamtsystem aus vielen Teilchen besteht. Dies ist bei allen makroskopischen Systemen erfüllt. In der Wärmelehre werden vorwiegend Gleichgewichtszustände betrachtet. Der *thermodynamische Gleichgewichtszustand* stellt sich ein, wenn man die äußeren, aufgeprägten Bedingungen, Druck oder Temperatur, zeitlich nicht ändert. In welchem Zeitraum der Gleichgewichtszustand erreicht wird, ist sehr unterschiedlich. Er hängt davon ab, wie schnell die Prozesse ablaufen, über die sich das System dem Gleichgewichtszustand nähert. In Festkörpern bleiben daher metastabile Zustände, z. B. die amorphe Struktur, oftmals eingefroren.
Nichtgleichgewichtszustände werden bei Transportprozessen (Thermodynamik irreversibler Prozesse) und bei Schwankungserscheinungen (Nichtgleichgewichtsstatistik) untersucht.

2.6.1. Grundbegriffe

2.6.1.1. Stofflicher Aufbau der Materie. Die Substanz, die in der Wärmelehre unser physikalisches System darstellt, ist durch ihre *chemische Zusammensetzung* – welche Elemente, welche Moleküle sind in welchen Konzentrationen enthalten – charakterisiert. Weiterhin kann die Masse oder Menge angegeben werden, das Volumen, welches sie einnimmt, oder ihre Dichte. Wenn nicht explizit etwas anderes zum Ausdruck gebracht wird, werden in der Wärmelehre homogene Systeme mit konstanter Dichte betrachtet – auch für die einzelnen Komponenten. Als *Aggregatzustände* werden fest, flüssig und gasförmig unterschieden.
Festkörper haben ein gegebenes Volumen und eine feste Form. Die Volumenänderungen bei Erhöhung des äußeren Druckes sind relativ klein. Für eine Verformung sind relativ große Kräfte erforderlich. Bis zu einer gewissen Stärke der äußeren Einwirkung stellt sich nach deren Wegnahme die ursprüngliche Form wieder ein.
Flüssigkeiten haben ein gegebenes Volumen, aber keine feste Form. Volumenänderungen bei Erhöhung des äußeren Druckes sind wie beim Festkörper klein. Flüssigkeiten passen sich dem Gefäß an, in dem sie aufbewahrt werden, d. h. eine Formänderung ist leicht herbeiführbar. In den Grenzfällen zum Festkörper hin hat die Flüssigkeit eine große Zähigkeit.
Gase in der *Dampfphase* füllen immer das gesamte, ihnen zur Verfügung stehende Volumen aus. Um mit wohldefinierten Gasmengen zu arbeiten, benötigt man daher in der Regel geschlossene Behälter. Ein Gas ist relativ leicht komprimierbar.
Komponenten. Besteht das betrachtete System nur aus einem Element oder einer chemischen Verbindung, dann spricht man von einer *reinen Substanz*. *Gemische* enthalten verschiedene Komponenten. Luft besteht im wesentlichen aus Stickstoff (78 %) und Sauerstoff (21 %), Messing aus Kupfer und Zink.
Phasen nennt man verschiedene Teile des Systems, die sich durch ihren Aggregatzustand oder ihre chemische Zusammensetzung voneinander unterscheiden. Im thermodynamischen Gleichgewicht können verschiedene Phasen nebeneinander bestehen. Über der flüssigen Phase Wasser befindet sich am Siedepunkt die Phase Wasserdampf, eine gesättigte Zuckerlösung steht mit Zuckerkristallen im Gleichgewicht.
2.6.1.2. Temperatur. Mit den menschlichen Sinnesempfindungen können warm und kalt unterschieden werden. Ein System kann wärmer als ein anderes sein. Die Temperatur ist der physikalische Begriff, mit dem diese Eigenschaft klassifiziert wird. Sie wird mit T in der thermodynamischen Skala und mit ϑ in der Celsius-Skala bezeichnet. Die genaue Festlegung der Temperatur verlangt eine Meßvorschrift und die Definition einer Einheit. Dies geschieht über die *thermodynamische Temperaturskale*. Die Temperatur ist eine intensive Größe.
Bei Temperaturmessungen im täglichen Leben wird die Celsius-Skala verwendet. Bei wissenschaftlichen Betrachtungen wird die thermodynamische Temperatur (oder *absolute Temperatur*) bevorzugt.
Die *Celsius-Skala* (ϑ) geht von zwei Fixpunkten aus. Die Temperatur des Eispunktes von Wasser bei Normaldruck, d. h. die Temperatur, bei der Eis und mit Luft gesättigtes Wasser miteinander im thermodynamischen Gleichgewicht stehen, wird als 0 °C (Grad Celsius) bezeichnet. Die Temperatur, bei der Wasser und sein Dampf unter Normaldruck miteinander im Gleichgewicht stehen, wird zu 100 °C festgelegt. Zustände, die unter 0 °C liegen, erhalten negative Temperaturen.
Temperaturmessung. Hierbei wird die Temperaturabhängigkeit physikalischer Größen, insbesondere des Volumens und des elektrischen Widerstandes, ausgenutzt. Verbreitetstes Meßinstrument zur Temperaturmessung ist das *Thermometer*.
Flüssigkeitsthermometer. Da sich die Flüssigkeit stärker ausdehnt als das Glasröhrchen, in dem sie sich befindet, steigt

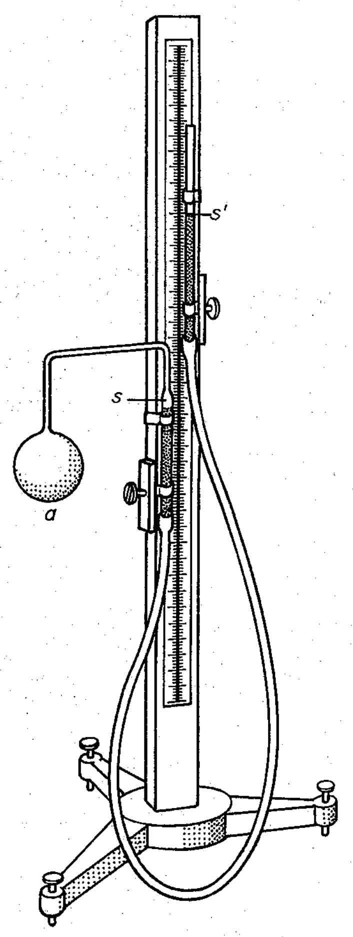


Abb. 2.6.1–1 Jollysches Luftthermometer

Tab. 2.6.1–2 Fixpunkte der Internationalen Praktischen Temperaturskala von 1968 bezogen auf 101 325 Nm^{-2}

Tripelpunkt von Wasserstoff[1]	13,81 K
Siedetemperatur von Wasserstoff[1] bei 33 306,6 Nm^{-2}	17,042 K
Siedepunkt von Wasserstoff[1]	20,28 K
Siedepunkt von Neon	27,102 K
Tripelpunkt von Sauerstoff	54,361 K
Siedepunkt von Sauerstoff	90,188 K
Tripelpunkt von Wasser[2] und[3]	273,16 K
Siedepunkt von Wasser[3]	373,15 K
Erstarrungspunkt von Zinn	505,1181 K
Erstarrungspunkt von Zink	692,73 K
Erstarrungspunkt von Silber	1 235,08 K
Erstarrungspunkt von Gold	1 337,58 K

[1] im Ortho-/Para-Wasserstoff-Gleichgewicht, [2] genau per Definition, [3] Wasserstoff- und Sauerstoffisotope wie im Ozeanwasser

beim Erwärmen der Flüssigkeitsspiegel. Im Thermometer befinden sich nur kleine Flüssigkeitsmengen, damit die Flüssigkeit rasch die Temperatur der Umgebung annimmt und die von der Flüssigkeit aufgenommene Wärmeenergie die Temperatur der untersuchten Probe nicht wesentlich verändert. Das Röhrchen ist geschlossen, damit nicht durch Verdunsten der Flüssigkeit eine Temperaturabnahme vorgetäuscht wird. Je nach der Flüssigkeitsmenge und dem Durchmesser der Kapillaren, in der die Flüssigkeit steigt, erhält man bei gleicher Temperaturerhöhung verschiedene Steighöhen. Je dünner die Kapillare ist, desto kleinere Temperaturunterschiede sind ablesbar, wobei natürlich entsprechende Forderungen an die Güte der Kalibrierung zu stellen sind. Flüssigkeitsthermometer sind nur bis zum Gefrierpunkt der Flüssigkeit einsetzbar. Für Quecksilber beträgt dieser −39 °C. Bis −100 °C können Alkohol und Toluol verwendet werden, bis zu −200 °C Pentan. Die Volumenausdehnung der Flüssigkeit ist nicht genau proportional zur Temperaturerhöhung. Bei Thermometern, die hohe Genauigkeitsansprüche erfüllen, sind daher die Skalenstriche nicht genau äquidistant. Außerdem sind diese Abweichungen für verschiedene Flüssigkeiten unterschiedlich. Daher decken sich die Einteilungen zweier Thermometer, die verschiedene Flüssigkeiten enthalten, auch dann nicht, wenn die Abstände zwischen den Marken für 0 °C und 100 °C gleich gewählt wurden.

Gasthermometer sind gut zum Eichen von Thermometern geeignet. Die Zustandsgleichung des idealen Gases lautet pV/T = konst. Die Temperaturmessung kann daher auf eine Volumen- und Druckmessung zurückgeführt werden. Bei konstantem Volumen wächst der Druck proportional zur Temperatur an, $p(T) = p(T_0) + a(T - T_0)$. Die in der Zustandsgleichung des idealen Gases auftretende Temperatur stimmt mit der thermodynamischen Temperatur überein. Reale Gase verhalten sich weit oberhalb des Siedepunktes – z. B. Wasserstoff bei Zimmertemperatur – wie ideale Gase. Abb. 2.6.1–1 zeigt das *Jollysche Luftthermometer*. Ein mit trockener Luft gefülltes Gasvolumen (a) steht unter dem Druck, den Atmosphäre und eine Quecksilbersäule ($s' - s$) bewirken. Durch Veränderung der Höhe der Quecksilbersäule (s') kann das Volumen konstant gehalten werden (feste Marke s). Umgibt man das Gasvolumen mit schmelzendem Eis, dann kann man den zu 0 °C gehörenden Druck ablesen. Bringt man das Gasvolumen in siedendes Wasser, dann erhält man den zu 100 °C gehörenden Druck.

Die *thermodynamische Temperatur T* ist über den Wirkungsgrad eines reversiblen Kreisprozesses definiert. Temperatur wie auch Temparturdifferenzen werden in *Kelvin* angegeben. Die Einheit Kelvin wurde dadurch festgelegt, daß dem Tripelpunkt des Wassers die Temperatur 273,16 K zugeordnet wurde. Ein Kelvin ist der 273,16te Teil der absoluten Temperatur am Tripelpunkt des Wassers. Der gebrochene Zahlenwert rührt daher, daß man die Kelvin-Skala – was Temperaturdifferenzen betrifft – möglichst gut mit der Celsius-Skala in Übereinstimmung bringen wollte. $\vartheta = 0$ °C entspricht $T = 273{,}15\,K$, da der Eispunkt des Wassers 0,01 K unter dem Tripelpunkt liegt. Die Temperaturdifferenz 1 °C ist gleich der Temperaturdifferenz 1 K. Während man Kelvin wie andere Einheiten mit Vorsätzen, z. B. Millikelvin (mK), versehen kann, ist dies für die Einheit °C nicht erlaubt.

Die Temperatur ist – mikroskopisch gesehen – ein Maß für die Wärmebewegung der Atome. Ist deren kinetische Energie auf ihren untersten Wert reduziert worden – die Quan-

tenphysik lehrt, daß dieser nicht Null ist –, kann die Substanz nicht weiter abgekühlt werden. Der absolute Nullpunkt, der Nullpunkt 0 K der thermodynamischen Temperaturskala, ist erreicht worden. Bezüglich der Celsius-Skala liegt er bei $\vartheta = -273{,}15$ °C. Der Begriff negative absolute Temperaturen wird in einem anderen Sinne verwendet. Am geeignetsten für die Messung absoluter Temperaturen sind Experimente, die über die Zustandsgleichung des idealen Gases bzw. über das Plancksche Strahlungsgesetz die Temperaturmessung auf eine Volumen- und Druckmessung bzw. eine Messung der Strahlungsleistung zurückführen. Diese Messungen sind jedoch sehr aufwendig und daher schwer reproduzierbar. Aus diesem Grund wurden bestimmte *Fixpunkte* (Tab. 2.6.1-2) ausgemessen. Zwischen diesen Fixpunkten ist eine Messung oder Eichung mit speziellen Widerstandsthermometern bzw. Thermoelementen möglich.

2.6.1.3. Wärmeenergie. Ein Körper kann erwärmt oder gekühlt werden, indem eine Wärmemenge zu- oder abgeführt wird. Der Begriff „Wärmemenge" erinnert noch daran, daß man Wärme zuerst als einen besonderen Stoff ansah, der beim Erwärmen oder Abkühlen übertragen wird. Für die Wärmeenergie hatte man eine eigene Einheit, die *Kalorie* (cal) eingeführt. Heute werden Wärmeenergien in der gleichen Einheit gemessen wie andere Energieformen (Joule). Die Wärmeenergie ΔQ, die man benötigt, um die Temperatur eines Körpers um ΔT zu erhöhen, hängt von der Art und der Masse des Körpers ab. Als *Wärmekapazität* bezeichnet man den Quotienten $\Delta Q/\Delta T$ aus zugeführter Wärmemenge und erzielter Temperaturerhöhung. Die Wärmeenergie ist eine extensive Größe. Daher bezieht man die Wärmekapazität auf ein Mol (*Molwärme C*) bzw. man verwendet die *spezifische Wärme c* (vgl. Tab. 2.6.5–3), $\Delta Q = zC\Delta T = Mc\Delta T$ (z = Molzahl). Die spezifische Wärme ist temperaturabhängig. Die Definition der Kalorie war mit der spezifischen Wärme des Wassers verknüpft. Die Wärmemenge, die man benötigt, um 1 g Wasser unter Normaldruck von 14,5 auf 15,5 °C zu erwärmen, bezeichnete man als 1 cal.

Beim Temperaturausgleich zwischen zwei Systemen, z. B. beim Mischen zweier Wassermengen der Masse M_1 bzw. M_2 und der Temperatur ϑ_1 bzw. ϑ_2, nimmt das kältere System soviel Wärmeenergie auf, wie das wärmere System abgibt, Wärmeisolation gegenüber anderen Systemen vorausgesetzt. Die Mischtemperatur $\vartheta_m = (M_1 c_1 \vartheta_1 + M_2 c_2 \vartheta_2)/(M_1 c_1 + M_2 c_2)$ kann aus dieser Bilanz berechnet werden.

2.6.1.4. Arbeit bei Ausdehnung und Kompression. Druck ist eine abgeleitete Größe, seine Einheit ist *Pascal* (1 Pa=1 N/m², Tab. 2.6.1–3). Aus mikroskopischer Sicht entsteht der Druck dadurch, daß die Atome oder Moleküle eines Gases mit der Geschwindigkeit, die sie infolge der Wärmebewegung haben, auf die Wand auftreffen und reflektiert werden. Die dabei auftretende Impulsänderung entspricht einer Kraftwirkung. Ein Gas kann andererseits auf einen bestimmten Druck gebracht werden, indem über einen Stempel mit der Fläche A (Abb. 2.6.1–4) eine Kraft F, z. B. durch die Schwerkraft einer Masse M, übertragen wird. Der sich im thermodynamischen Gleichgewicht einstellende Druck ist $p = F/A$.

Erhöht man die Kraft ($\mathrm{d}M$) und damit den Druck, so stellt sich ein neuer Gleichgewichtszustand ein (Abb. 2.6.1–4b). Bewegt sich der Stempel um eine Strecke $\mathrm{d}x$, dann hat sich das Volumen des Gases um $\mathrm{d}V = A\,\mathrm{d}x$ verringert. Bei dieser Kompression hat die Kraft F die Arbeit $W = -F\mathrm{d}x$ geleistet ($\mathrm{d}x$ und $\mathrm{d}V$ sind hier negativ), die auch $W = -p\,\mathrm{d}V$ geschrieben werden kann. Die dem Gas bei einer Volumenänderung von V_1 auf V_2 zugeführte Arbeit ist also

$$W_{\text{zugef.}} = -\int_{V_1}^{V_2} p\,\mathrm{d}V.$$

Bei einer Expansion ($V_2 > V_1$) leistet das Gas Arbeit, $W_{\text{zugef.}} < 0$. Im Beispiel der Abb. 2.6.1–4 wird diese Arbeit genutzt, um die potentielle Energie der Masse M zu erhöhen.

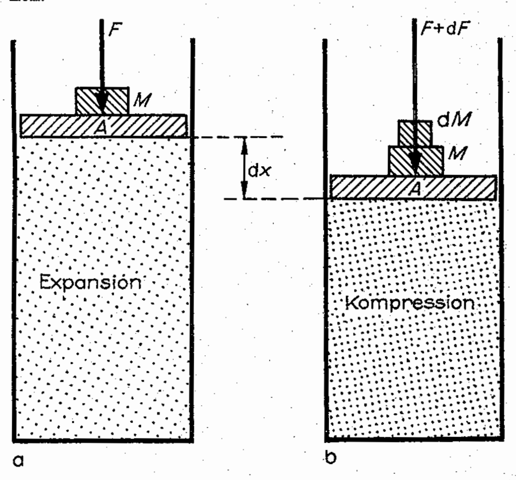


Abb. 2.6.1–4 Kompression und Expansion

Tab. 2.6.1–3 Druckeinheiten

Einheit	Symbol	Bedeutung
Pascal	Pa	$1\ \text{Pa} = 1\ \text{Nm}^{-2}$
Millibar	mbar	1 mbar = 100 Pa
Technische Atmosphäre[1]	at	$1\ \text{at} = 0{,}980665 \cdot 10^5\ \text{Pa}$
Torr[2]	Torr	1 Torr = 133,3 Pa
Physikalische Atmosphäre[3]	atm	1 atm = 760 Torr $= 1{,}01325 \cdot 10^5\ \text{Pa}$
Meter Wassersäule	mWS	1 mWS = 0,1 at

[1] 1 at = 1 kp/cm², [2] 1 Torr = 1 mm Quecksilbersäule,
[3] 1 atm = mittlerer Luftdruck = Normaldruck p_0

2.6.1.5. Zustand, Freiheitsgrade. Der *thermodynamische Gleichgewichtszustand*, kurz Zustand genannt, ist dadurch bestimmt, daß keine Wärmezufuhr mehr aus der Umgebung auftritt und das System eine einheitliche Temperatur aufweist, daß keine Volumen- und Druckänderung mehr stattfindet und daß jede Phase des Systems homogen ist.

Zustandsgrößen nennt man die makroskopischen Parameter, die in diesem Zustand eindeutig festliegen, vor allem also nicht davon abhängen, wie das System in diesen Zustand gekommen ist. Die Zustandsgrößen können durch die äußeren Bedingungen festgelegt sein, wie Temperatur, Volumen oder Druck, es können aber auch für das betrachtete System charakteristische Größen sein, wie innere Energie, Kompressibilität usw. Eine bestimmte Anzahl von Zustandsgrößen kann man vorgeben, dem System aufprägen. Diese Anzahl nennt man die Zahl der Freiheitsgrade des Systems. Die anderen Zustandsgrößen sind eindeutige Funktionen dieser herausgegriffenen Zustandsgrößen.

In den einfachsten Beispielen hat das System zwei *Freiheitsgrade*. Man kann ein Gas in ein Gefäß mit festem Volumen einschließen, d.h. das Volumen vorgeben, und dieses Gefäß in ein Wärmebad stellen, so daß das Gas die Temperatur des Wärmebades annimmt. Also sind Temperatur und Volumen frei vorgebbar. Der Wert des Druckes stellt sich dabei als Funktion von T und V ein. Man kann auch den Wert des Druckes vorgeben, dann stellt sich ein durch Temperatur und Druck bestimmtes Volumen ein. Die Zahl der Freiheitsgrade erhöht sich bei der Betrachtung von Systemen aus mehreren *Komponenten*. Die *Konzentrationen* sind frei wählbar und damit weitere Freiheitsgrade. Die Gesamtmasse wird dagegen nicht als Freiheitsgrad gezählt, obwohl sie auch frei wählbar ist. Aber eine Verdopplung aller extensiven Größen bringt physikalisch nichts Neues, sie ist der Betrachtung zweier Systeme der Ausgangsmenge äquivalent. Befindet sich das betrachtete System in *äußeren Feldern*, z. B. in einem Magnetfeld, dann erhöht sich die Anzahl der Freiheitsgrade, da jede Feldstärke frei wählbar ist. Durch Nebenbedingungen kann die Anzahl der Freiheitsgrade eingeschränkt werden. Wird z. B. verlangt, daß Dampfphase und Kondensat miteinander im Gleichgewicht stehen, dann gibt es bei vorgegebenem Druck nur eine Temperatur, bei der dies realisierbar ist, bei Wasser unter Normaldruck z. B. nur die Temperatur 100 °C. Eine allgemeingültige Aussage über die Zahl der Freiheitsgrade macht die *Gibbssche Phasenregel* (vgl. 2.6.6.4.). Welche Zustandsgrößen man entsprechend der Zahl der Freiheitsgrade als Parameter auswählt, wird durch Zweckmäßigkeitsgründe bei der jeweiligen Aufgabenstellung bestimmt. Es können z. B. T und V oder T und p o. a. Kombinationen sein.

2.6.1.6. Extensive und intensive Größen. Extensive Größen eines Systems sind solche, die proportional zur Masse oder Teilchenzahl sind. Masse und Teilchenzahl selbst sind damit auch extensive Größen. Als weitere Beispiele sind Volumen, innere Energie und Entropie zu nennen.

Intensive Größen sind unabhängig von der Menge, z. B. Temperatur und Druck. Der Quotient zweier extensiver Größen ist auch eine intensive Größe, so die Dichte als Quotient aus Masse und Volumen oder die Konzentration als Quotient aus der Teilchenzahl einer Komponente und der Gesamtteilchenzahl.

Mengeneinheit. Viele Aussagen der Wärmelehre sind unabhängig von der Größe des Systems, so die Lage des Schmelz, Siede- oder Tripelpunktes wie auch der Wirkungsgrad eines Kreisprozesses. Man hat daher eine Mengeneinheit eingeführt, auf die man alle extensiven Größen bezieht. Die Mengeneinheit ist das *Mol* (mol). 1 mol einer Substanz sind soviel Gramm, wie die relative Atom- oder Molekülmasse angibt. Damit ist der Molbegriff an eine Komponente gebunden. Für ein Gemisch muß man die Molzahl der einzelnen Komponenten angeben. Die Definition der Mengeneinheit wurde so gewählt, daß sich für alle Substanzen in der Menge 1 mol die gleiche Teilchenzahl befindet. Diese Anzahl ist durch die *Avogadro-Konstante* gegeben, $L = 6{,}0221 \cdot 10^{23}\,\mathrm{mol}^{-1}$, auch *Loschmidtsche Konstante* oder *Loschmidtsche Zahl* genannt.

Da die relative Atommasse des ^{12}C-Isotopes definitionsgemäß $A = 12$ ist, hat die Menge 1 mol ^{12}C die Masse 12 g. 30 g dieses Isotopes sind also 2,5 mol. Die Anzahl der Mole (hier 2,5) wird *Molzahl* (z) genannt. Die Gesamtmasse eines Systems ist also auch als Molzahl mal Molekulargewicht mal Gramm schreibbar, die Gesamtteilchenzahl als Molzahl mal Loschmidtsche Zahl, $N = zL$. Bei Gasen wie Sauerstoff und Stickstoff muß man beachten, daß sie aus Molekülen (O_2, N_2) bestehen. Ein Mol Sauerstoff ($A = 16{,}00$) sind also 32,00 g. Edelgase bestehen dagegen aus Atomen. Ein Mol Argon ($A = 39{,}994$) sind 39,994 g.

2.6.1.7. Zustandsgleichungen. Ein thermodynamisches System hat in den einfachsten Beispielen zwei Freiheitsgrade, man kann ihm z.B. die Werte für Volumen und Temperatur aufprägen. Zu diesen Werten V und T gehört im thermodynamischen Gleichgewicht ein bestimmter Druck p. Diesen Zusammenhang $p = p(V, T)$ nennt man die Zustandsgleichung, genauer *thermische Zustandsgleichung*, des Systems.

Das einfachste und zugleich auch wichtigste Beispiel ist die *Zustandsgleichung des idealen Gases*,

$$pV = zRT, \quad R = 8\,314{,}3\ \mathrm{J\,kmol^{-1}K^{-1}}.$$

Sie wurde experimentell an realen Gasen gefunden, wenn diese weit oberhalb des Kondensationspunktes untersucht wurden. z ist die Molzahl und R wird als *universelle Gaskonstante* bezeichnet. Wählt man T = konst., dann enthält die Zustandsgleichung des idealen Gases als Spezialfall das *Boyle-Mariottesche Gesetz* pV = konst. Hält man p = konst., dann erhält man für die Temperaturabhängigkeit des Volumens $V = V_0[1 + (T - T_0)/T_0]$, das *Gay-Lussac-Gesetz*. Wählt man als V_0 das Volumen bei 0 °C, dann findet man T_0 experimentell aus dem Ausdehnungskoeffizienten zu $T_0 = 273{,}15$ Grad. Deswegen wurde bei der Definition der Kelvin-Skala 0 °C die Temperatur 273,15 K zugeordnet. Geht man von der gleichen Stoffmenge (z. B. 1 mol) aus, dann genügen alle Gase der gleichen Zustandsgleichung, soweit diese als ideale Gase angesehen werden können. Das heißt z. B., daß unter *Normalbedingungen* ($p = 101\,325$ Pa und $\vartheta = 0\,°\mathrm{C}$) ein Mol eines Gases das gleiche Volumen (*Molvolumen* $V_m = 22{,}414$ l/mol) einnimmt.

In der Nähe des Kondensationspunktes beschreibt die Zustandsgleichung des idealen Gases reale Gase falsch. Als Modell einer Zustandsgleichung realer Gase wird die *van-der-Waalssche Zustandsgleichung* betrachtet,

$$[p - z^2a^2/V^2]\,(V - zb) = zRT.$$

Der Term $(az/V)^2$ beschreibt eine Druckverminderung durch die Anziehung zwischen den Atomen oder Molekülen, zb charakterisiert das Eigenvolumen der Atome.

Die van-der-Waalssche Zustandsgleichung (Abb. 2.6.1–5) weicht nur für geringe Dichten und hohe Temperaturen wenig von der des idealen Gases ab. Die Kurven T = konst. sind Kurven dritten Grades mit einem Wendepunkt. Die Temperatur, bei der die Tangente im Wendepunkt waagerecht verläuft, wird als kritische Temperatur T_{kr} bezeichnet. Bei darunterliegenden Temperaturen enthält die Kurve $p(V)$ ein Stück mit positivem Anstieg. Das ist unphysikalisch, da mit steigendem Druck das Volumen zunehmen würde. In diesem Bereich ist die Zustandsgleichung eine Gerade, auf der Flüssigkeit und Dampfphase miteinander im Gleichgewicht stehen (vgl. 2.6.7.3.). Die Gerade ist so zu legen, daß die beiden abgeschnittenen Flächen den gleichen Flächeninhalt haben. Das Modell van-der-Waals-Zustandsgleichung enthält also das Phänomen Kondensation. Sie wird daher oft für qualitative Betrachtungen herangezogen.

Nicht nur der Druck, auch andere Zustandsgrößen sind Funktionen von Volumen und Temperatur, so die spezifische Wärme, die Kompressibilität. Den Zusammenhang der inneren Energie $U = U(V, T)$ mit Volumen und Temperatur bezeichnet man als *kalorische Zustandsgleichung*.

2.6.1.8. Zustandsänderungen. Bei einem thermodynami-

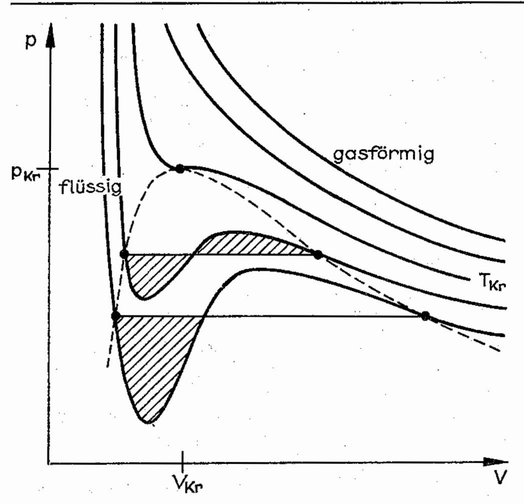


Abb. 2.6.1–5 Isothermen der Van-der-Waals-Zustandsgleichung

schen System mit zwei Freiheitsgraden kann man den Zusammenhang zwischen Druck, Volumen und Temperatur gut mittels des *Zustandsdiagrammes* veranschaulichen. Im p-V-Diagramm sind für das ideale Gas Linien konstanter Temperatur Hyperbeln $p \sim V^{-1}$. Natürlich kann man auch p und T oder T und V als Variable wählen. Ein thermodynamischer Gleichgewichtszustand wird im Zustandsdiagramm durch einen Punkt gekennzeichnet. Eine Zustandsänderung in einen anderen Gleichgewichtszustand führt zu einem anderen Punkt des Zustandsdiagrammes.
Man führt den Begriff der *quasistatischen Zustandsänderung* ein. Bei dieser sollen sich die Zustandsgrößen, insbesondere Druck und Temperatur, nur in infinitesimal kleinen Schritten ändern, und nach jedem Schritt soll sich das System im thermodynamischen Gleichgewicht befinden. Bei einer quasistatischen Zustandsänderung werden also nur Gleichgewichtszustände durchlaufen. Daher kann eine quasistatische Zustandsänderung durch eine Kurve im Zustandsdiagramm dargestellt werden.
Bei nichtquasistatischer Führung werden zwischenzeitlich Nichtgleichgewichtszustände angenommen. So wird in Abb. 2.6.1–6 beim plötzlichen Wegnehmen der Trennwand durch die endliche Druckdifferenz zwischen gasgefülltem Teil und Vakuum ein Strömen des Gases bewirkt. Erst nach einer gewissen Zeit ist die Dichte konstant und die Bewegung der Atome reine ungeordnete Wärmebewegung. Die Nichtgleichgewichtszustände lassen sich nicht durch Punkte im Zustandsdiagramm beschreiben, sie sind durch die Werte von Druck und Temperatur nicht vollständig charakterisiert. Bei einer Zustandsänderung zwischen zwei Gleichgewichtszuständen ist die Änderung der Zustandsgrößen unabhängig davon, ob die Zustandsänderung quasistatisch oder nicht geführt wurde. Die Zustandsgrößen sind eindeutige Funktionen der Zustandsvariablen, also z.B. von Volumen und Temperatur. Die Änderung von Zustandsgrößen kann man daher immer berechnen, indem man sich eine zweckmäßige Zustandsänderung zwischen den entsprechenden Punkten heraussucht.

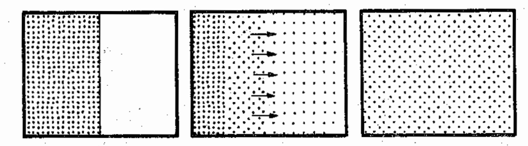


Abb. 2.6.1–6 Durchlaufen von Nichtgleichgewichtszuständen

Eine *isotherme Zustandsänderung* ist eine Zustandsänderung bei konstanter Temperatur. Im Zustandsdiagramm verläuft sie entlang der Isothermen T = konst. Bei einer *isobaren Zustandsänderung* gilt p = konst., bei einer *isochoren* V = konst. Bei einer *adiabatischen Zustandsänderung* wird dem System keine Wärmeenergie zugeführt, es ist also wärmeisoliert.
Als einen *Kreisprozeß* bezeichnet man eine Kombination von Zustandsänderungen, die zum Ausgangszustand zurückführt. Im Zustandsdiagramm wird ein Kreisprozeß durch eine geschlossene Kurve dargestellt.

2.6.2. Erster Hauptsatz der Wärmelehre

Der erste Hauptsatz stellt im wesentlichen die Aussage der Energieerhaltung bzw. eine Bilanzgleichung für die Energie dar. Diese Aussagen sind aus anderen Gebieten, insbesondere aus der Mechanik, bekannt. Es geht hier um die für die Wärmelehre spezifische Formulierung unter Einbeziehung der Wärmeenergie.
2.6.2.1. Wärme als Energieform. Es war eine bahnbrechende Erkenntnis, daß Wärme eine Energieform ist, daß damit auch Wärmeenergie in mechanische Energie umwandelbar sein muß. Sie bildete die Grundlage für den Bau von Dampfmaschinen. ROBERT MAYER gab als erster das *mechanische Wärmeäquivalent* an,

1 cal = 0,4279 kpm = 4,1868 Joule.

Diese Äquivalenz ist so trivial geworden, daß auf eine eigene Einheit für die Wärmeenergie verzichtet wird. Die Einheit Kalorie ist nicht mehr zu verwenden. Auch Wärmeenergien werden in Joule (Nm) gemessen.

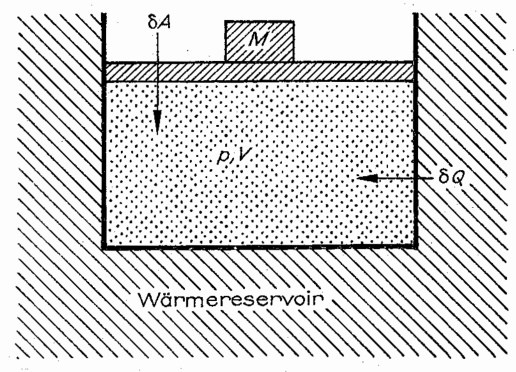


Abb. 2.6.2–1 Zufuhr von Wärme und Arbeit

2.6.2.2. Innere Energie. Der Energieinhalt eines Systems wird in der Wärmelehre als innere Energie U bezeichnet. Sie kann verändert werden, indem Energie zu- oder abgeführt wird. In der Wärmelehre interessieren vorrangig Energieänderungen durch Zuführen (oder Abführen) von Wärmeenergie und Arbeit. Bei Anwesenheit von elektrischen oder magnetischen Feldern ist natürlich die elektromagnetische Energie einzubeziehen.

Die *innere Energie* kann erhöht werden ($\delta U > 0$), indem man *Wärme* δQ oder *Arbeit* δW zuführt, $\delta U = \delta Q + \delta W$ (Abb. 2.6.2–1). In diese Aussage ist die Erkenntnis, daß Wärme eine Energieform ist, eingearbeitet. Man beachte die Vorzeichenfestlegung.

δQ bzw. δW sind positiv, wenn dem System Wärme bzw. Arbeit zugeführt wird, negativ, wenn dem System Wärmeenergie entzogen wird bzw. wenn das System Arbeit leistet:

$\delta Q > 0$	Wärmeenergie wird zugeführt,
$\delta Q < 0$ bzw. $\delta\overline{Q} = -\delta Q > 0$	Wärmeenergie wird abgeführt,
$\delta W > 0$	Arbeit wird zugeführt,
$\delta W < 0$ bzw. $\delta\overline{W} = -\delta W > 0$	Arbeit wird abgegeben.

Die Arbeit δW hängt über $\delta W = -p\,dV$ mit der Volumenänderung dV und dem Druck p zusammen (vgl. 2.6.1.4.). Bei einer Kompression $V_2 \to V_1$ ist die zugeführte Arbeit $W = -\int_{V_2}^{V_1} p\,dV = \int_{V_1}^{V_2} p\,dV$, im Zustandsdiagramm also die Fläche unter der Kurve, die die Zustandsänderung beschreibt (Abb. 2.6.2–2a). Bei einem Kreisprozeß ist die Arbeit gleich der eingeschlossenen Fläche (Abb. 2.6.2–2b). Sie ist positiv ($W > 0$), d.h. dem System wurde Arbeit zugeführt, wenn die Zustandsänderungen die Fläche im mathematisch positiven Sinne umlaufen. Wird der Kreisprozeß im Uhrzeigersinn durchlaufen, dann ist W negativ, und das System gibt Arbeit ab.

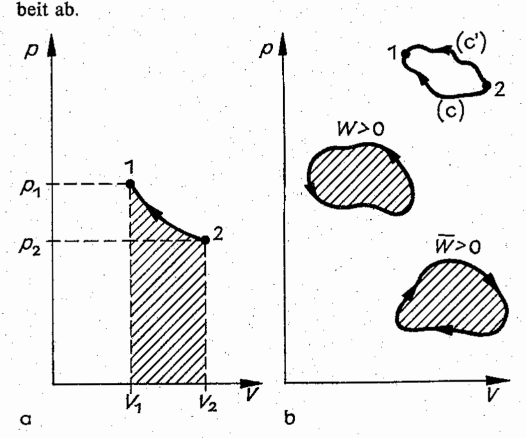


Abb. 2.6.2–2 Arbeit (a) als Fläche im Zustandsdiagramm, (b) beim Kreisprozeß

2.6.2.3. Formulierungen des ersten Hauptsatzes. Neben der Feststellung, daß Wärme eine Energieform ist, enthält der erste Hauptsatz die Aussage: *Die innere Energie ist eine Zustandsgröße*. Sie ist damit eine eindeutige Funktion der Zustandsvariablen, z. B. von Volumen und Temperatur, $U = U(V, T)$. Mathematisch wird diese Aussage dadurch zum Ausdruck gebracht, daß eine Änderung δU durch ein totales Differential dU beschrieben wird, $dU = \delta Q - p\,dV$. Nach Durchlaufen eines Kreisprozesses gelangt das betrachtete System in seinen Ausgangszustand zurück. Da die innere Energie eine Zustandsgröße ist, hat sie den gleichen Wert wie zu Beginn des Kreisprozesses. Hat das System z. B. während des Kreisprozesses Arbeit geleistet, so muß ihm nach dem ersten Hauptsatz eine entsprechend große Wärmemenge zugeführt worden sein. Daraus leitet sich eine andere Formulierung des ersten Hauptsatzes ab: Es gibt kein *Perpetuum mobile erster Art*. Darunter versteht man eine Maschine, die periodisch arbeitet, in jeder Periode Energie abgibt, aber sonst seine Umgebung nicht verändert. Bevor der Satz von der Erhaltung der Energie Allgemeingut geworden war, wurde versucht, ein perpetuum mobile zu konstruieren. Ein Automotor arbeitet periodisch und gibt auch in jeder Periode Energie (als Arbeit und Wärme) ab. Aber er verändert zusätzlich seine Umgebung, indem er Kraftstoff verbraucht.

Mathematisch äquivalente Formulierungen des ersten Hauptsatzes sind die Aussagen

$$\oint dU = 0 \text{ bzw. } \int_{2\,(c)}^{1} dU = \int_{2\,(c')}^{1} dU = U_1 - U_2.$$

Im ersten Integral wird ein Kreisprozeß betrachtet, über einen geschlossenen Weg im Zustandsdiagramm integriert. Die Änderungen der inneren Energie summieren sich zu Null auf, da die innere Energie zu Beginn und am Ende des Kreisprozesses gleich groß ist. Die zweite Gleichung bringt zum Ausdruck, daß die Änderung der inneren Energie bei einer Zustandsänderung vom Zustand 2 in den Zustand 1 (vgl. Abb. 2.6.2–1b) unabhängig davon ist, auf welchem Wege der Zustand 2 in den Zustand 1 überführt wird.

Arbeit und Wärme sind keine Zustandsgrößen. Es gibt also nicht „die Arbeit des Systems" oder „die Wärmeenergie des Systems". Da die Arbeit gleich der Fläche unter der Zustandsänderung ist (vgl. Abb. 2.6.2–1a), ist die bei der Kompression (vgl. Abb. 2.6.2–1b) zugeführte Arbeit für die Zustandsänderungen auf den Wegen C bzw. C' verschieden groß. Die dem System zugeführte Arbeit hängt also von der Vorgeschichte ab, ist keine eindeutige Funktion des erreichten Zustandes. Die gleiche Aussage gilt für die Wärmeenergie.

2.6.3. Zweiter Hauptsatz der Wärmelehre

Der zweite Hauptsatz ist erkenntnistheoretisch das Kernstück der Thermodynamik. Er hat keine äquivalenten Formulierungen in anderen Gebieten der Physik, wie es für den ersten Hauptsatz der Fall ist. Der zweite Hauptsatz ist eine Aussage über makroskopische Systeme, d. h. Systeme mit großer Teilchenzahl.

2.6.3.1. Unmöglichkeit eines Perpetuum mobile zweiter Art. Es gehört zu den prinzipiellen Erfahrungen der Wärmelehre, daß beim Wärmekontakt zweier Systeme unterschiedlicher Temperatur ein Temperaturausgleich stattfindet, daß Wärmeenergie vom wärmeren auf das kältere System übertragen wird. Es tritt nie auf, daß sich ein Teil eines Systems ohne äußere Eingriffe abkühlt und sich dafür ein anderer Teil erwärmt, also ein Temperaturunterschied entsteht, obwohl dies dem ersten Hauptsatz der Wärmelehre nicht widersprechen würde.

Eine äquivalente Formulierung dieses Sachverhaltes ist die Aussage: Es ist unmöglich, ein *Perpetuum mobile zweiter Art* zu konstruieren. Dies ist eine Maschine, die periodisch arbeitet, in jeder Periode Arbeit abgibt und dafür die gleiche Wärmemenge aus e i n e m Wärmespeicher aufnimmt. Gäbe es eine solche Maschine, dann könnten z.B. Schiffsmotoren angetrieben werden, indem das Meerwasser abgekühlt wird. Es ständen beliebig große Energiereserven zur Verfügung. Um das Wesentliche dieser Aussage zu erkennen, muß man sich vor Augen halten, daß Dampfmaschinen mit z w e i Wärmebehältern gekoppelt sind, einem heißen und einem kalten, wobei nur ein Teil der Wärmeenergie, die man dem

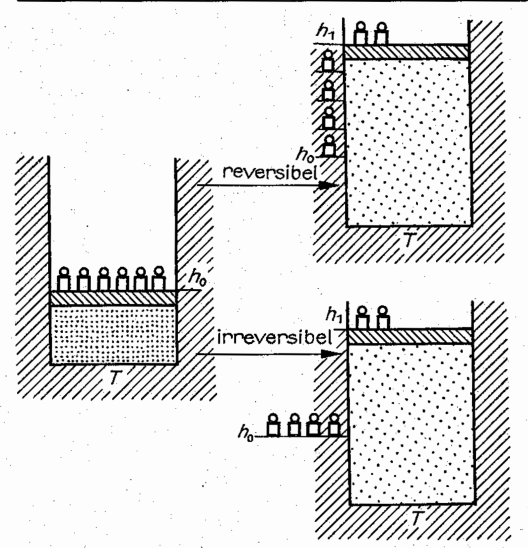


Abb. 2.6.3-1 Isotherme Expansion

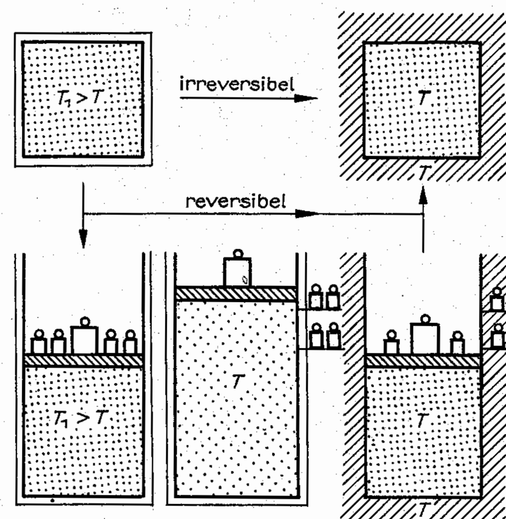


Abb. 2.6.3-2 Isochorer Temperaturausgleich

heißen Behälter entnimmt, in Arbeit umgewandelt wird, ein Teil wird ungenutzt an den kälteren Behälter abgegeben. Wärmeenergie kann – so lehrt die Erfahrung – nicht vollständig in Arbeit umgewandelt werden. Das Umgekehrte ist möglich, mechanische oder elektrische Energie können vollständig in Wärmeenergie überführt werden. Es gibt offensichtlich verschiedene Qualitäten von Energie. Der zweite Hauptsatz der Wärmelehre stellt eine quantitative Formulierung des diesen Erfahrungen zugrunde liegenden Gesetzes dar. Er gilt für endliche Systeme und kann nicht auf den Kosmos als Ganzes übertragen werden (vgl. 2.6.11.6.).

2.6.3.2. Reversible und irreversible Prozesse. Für die Formulierung des zweiten Hauptsatzes der Wärmelehre sind die Begriffsbildungen reversibel und irreversibel wichtig. Ein *reversibler Prozeß* kann auch in umgekehrter Richtung durchlaufen werden, wobei keine Veränderungen in der Umgebung zurückbleiben. Ist eine solche Umkehrung des Prozesses nicht möglich, ist es ein *irreversibler Prozeß*. Die nachfolgenden Beispiele zeigen, daß – vom gleichen Anfangszustand ausgehend – ein Endzustand sowohl durch reversible wie auch durch irreversible Prozeßführung erreicht werden kann.

Soll eine *isotherme Expansion* (Abb. 2.6.3-1) quasistatisch verlaufen, dann darf der Druck, den der Stempel entsprechend der angreifenden Kraft ausübt, nur infinitesimal unter dem Druck des Gases liegen. Die Kraft – in der Abbildung durch die Schwerkraft auf Massen bestimmt – muß also langsam verringert werden. Verringert man die Kraft auf den Stempel plötzlich um den gesamten Betrag, der der Druckdifferenz zwischen Anfangs- und Endzustand entspricht, dann dehnt sich das Gas explosionsartig aus. Es durchläuft Nichtgleichgewichtszustände, bis sich der gleiche Gleichgewichtszustand wie bei der quasistatischen Prozeßführung eingestellt hat. Der Unterschied in der Veränderung der Umgebung besteht darin, daß bei der quasistatischen Expansion mehr Arbeit abgegeben worden ist als im zweiten Fall. Bei der quasistatischen Führung verteilen sich die entfernten Massen auf den Bereich zwischen den Stempelhöhen h_0 und h_1, im zweiten Fall liegen sie alle auf der Höhe h_0, haben also eine kleinere potentielle Energie als im ersten Fall. Betrachtet man ein ideales Gas, bei dem die innere Energie vom Volumen unabhängig ist, dann ist die bei der Expansion geleistete Arbeit gleich der Wärmeenergie, die dem Wärmebad entzogen wurde. Bei der plötzlichen Expansion wurde also die Möglichkeit, Wärme in Arbeit umzuwandeln, nicht ausgeschöpft.

Die quasistatische Prozeßführung ist umkehrbar, reversibel. Man schiebt nur die Massestückchen nach und nach auf den Stempel, was ohne Arbeit möglich ist, um den Druck langsam wieder zu erhöhen. Die explosionsartige Expansion ist irreversibel. Man muß eine zusätzliche Energiequelle anzapfen, um die Massen auf die jeweilige Stempelhöhe anzuheben.

Abb. 2.6.3-2 veranschaulicht, wie ein *isochorer Temperatur-*

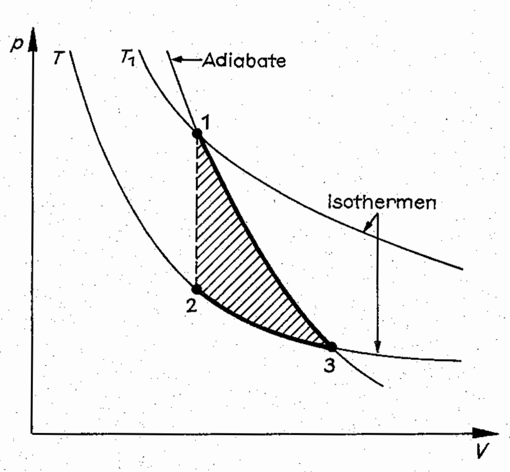


Abb. 2.6.3-3 Zustandsänderungen beim reversiblen, isochoren Temperaturausgleich

ausgleich reversibel geführt werden kann. Bei irreversibler Prozeßführung fließt einfach Wärme vom heißen Gas in das Wärmebad ab, bis das Gas dessen Temperatur T angenommen hat. Ohne äußere Einwirkung kann der Temperaturunterschied dann nicht wiederhergestellt werden. Bei der reversiblen Führung wird das System zunächst wärmeisoliert und dann durch eine adiabatische Expansion abgekühlt (Weg $1 \to 3$ in Abb. 2.6.3–3). Danach wird die Wärmeisolation entfernt und über eine isotherme Kompression (Weg $3 \to 2$) der Endzustand hergestellt. Da die bei der adiabatischen Expansion gewonnene Arbeit größer als die für die isotherme Kompression benötigte ist – der Unterschied entspricht der schraffierten Fläche in Abb. 2.6.3–3 – wurde bei der reversiblen Führung im Unterschied zur irreversiblen Führung Arbeit gewonnen und dafür weniger Wärmeenergie an das Wärmebad abgegeben.

Diese Beispiele zeigen, daß bei einer reversiblen Führung kein Wärmestrom zwischen Systemen mit einem endlichen Temperaturunterschied fließen darf (Wärmeleitung ist irreversibel) und daß keine endlichen Druckdifferenzen plötzlich ausgeglichen werden dürfen.

2.6.3.3. Carnotscher Kreisprozeß. Der Carnotsche Kreisprozeß ist ein reversibel geführter Kreisprozeß. Er ist besonders übersichtlich, da er aus zwei isothermen und zwei adiabatischen Prozessen besteht (Abb. 2.6.3–4). Zu seiner Führung sind zwei Wärmereservoire der Temperatur T_1 und T_2 ($T_1 > T_2$ angenommen) erforderlich.

Bei der isothermen Expansion $1 \to 2$ wird Arbeit geleistet und aus dem Wärmebehälter 1 die Wärmeenergie Q_1 entnommen. Bei der adiabatischen Expansion $2 \to 3$ wird Arbeit geleistet, die allein durch Verminderung der inneren Energie aufgebracht wird. Dabei kühlt sich das System auf die Temperatur T_2 ab. Bei der anschließenden isothermen Kompression $3 \to 4$ wird dem System Arbeit zugeführt. Das System selbst gibt die Wärmeenergie $\overline{Q}_2$ an das Wärmebad T_2 ab. Bei der adiabatischen Kompression $4 \to 1$ wird dem System Arbeit zugeführt. Diese wird allein zur Erhöhung der inneren Energie genutzt. Das System erwärmt sich auf die Temperatur T_1, die innere Energie hat ihren Ausgangswert wieder erreicht.

Die Veränderungen der Umgebung bei diesem Kreisprozeß bestehen darin, daß dem Wärmereservoir T_1 eine Wärmemenge Q_1 entnommen wurde, an das Wärmereservoir T_2 die Wärmeenergie $\overline{Q}_2$ abgegeben und Arbeit $\overline{W}$ geleistet wurde, die der Größe der umschlossenen Fläche entspricht. Aus dem ersten Hauptsatz der Wärmelehre folgt

$$\overline{W} = Q_1 - \overline{Q}_2.$$

Unter dem *Wirkungsgrad* η versteht man den Quotienten aus der gewonnenen Arbeit und der dem heißeren Wärmereservoir entnommenen Wärmeenergie,

$$\eta = \frac{\text{gewonnene Arbeit}}{\text{aufgewendete Wärmeenergie}} = \frac{Q_1 - \overline{Q}_2}{Q_1}.$$

Rechnet man die Teilprozesse des Carnotschen Kreisprozesses mit einem idealen Gas als Arbeitsmittel durch (vgl. 2.6.5.5.), dann ergibt sich für den Wirkungsgrad

$$\eta = (T_1 - T_2)/T_1.$$

Er ist durch T_1 und T_2 allein gegeben.

Aus der Aussage, daß kein perpetuum mobile zweiter Art konstruiert werden kann, folgt, daß der Wirkungsgrad aller reversiblen Kreisprozesse, die zwischen zwei Wärmereservoiren T_1 und T_2 arbeiten, gleich groß ist. Insbesondere ist der Wirkungsgrad unabhängig vom Arbeitsmittel, damit ist der Wirkungsgrad aller Kreisprozesse gleich dem für das ideale Gas berechnet. Gäbe es zwei reversible Kreisprozesse verschiedenen Wirkungsgrades η bzw. $\eta' < \eta$ (Abb. 2.6.3–5), dann müßte bei der Wahl $\overline{Q}_2 = \overline{Q}'_2$ also $Q'_1 < Q_1$ und damit $\overline{W}' < \overline{W}$ gelten. Koppelt man daher den Prozeß η mit dem in umgekehrter Richtung laufenden Prozeß η', dann wird insgesamt nur dem Wärmebehälter T_1 Wärmeenergie $Q_1 - Q'_1$ entzogen und dafür Arbeit $\overline{W} - \overline{W}'$ abgegeben. Das wäre aber ein Perpetuum mobile zweiter Art.

Für einen irreversiblen Prozeß ist der Wirkungsgrad kleiner als für einen reversiblen $\eta_{irr} < \eta_{rev}$, weil ja bei einem irreversiblen Prozeß in einem gewissen Maße die Möglichkeit vergeben wird, Wärme in Arbeit zu überführen.

2.6.3.4. Thermodynamische Temperaturskala. Da der Wirkungsgrad aller reversiblen Kreisprozesse zwischen zwei Wärmebehältern T_1 und T_2 gleich groß ist, kann er nicht von den besonderen Eigenschaften des Systems abhängen, er ist nur eine Funktion der beiden Temperaturen $\eta = 1 - \overline{Q}_2/Q_1 = 1 - f(T_2, T_1)$. Vergleicht man den Wirkungsgrad der beiden in Abb. 2.6.3–6 dargestellten Prozesse, dann findet man, daß sich $f(T_2, T_1) = g(T_2)/g(T_1)$ als Quotient darstellen lassen muß. Die darin auftretende Funktion der Temperatur $g(T)$ wächst mit der Temperatur monoton an, da $\eta = 1 - g(T_2)/g(T_1) > 0$ für $T_1 > T_2$ gilt. Einem Wärmebad ist also ein charakteristischer Parameter $g(T)$ zugeordnet, der über den Wirkungsgrad eines reversiblen Kreisprozesses definiert ist. Wie die Funktion $g(T)$ aussieht, hängt von der gewählten Temperaturskala ab. Bei der thermodynamischen Temperaturskala setzt man die Temperatur gleich dem Funktionswert $g(T)$. Dann ist der Wirkungsgrad eines reversiblen Kreisprozesses $\eta = 1 - T_2/T_1$. Da dieser Ausdruck mit demjenigen übereinstimmt, den man für das ideale Gas berechnet hat, ist die in der Zustandsgleichung des idealen Gases $pV = zRT$ auftretende Temperatur mit der thermodynamischen Temperatur identisch. Daher kann man mit dem Gasthermometer die thermodynamische Temperatur messen, solange man von dem Unterschied zwischen dem realen Gas und einem idealen Gas absehen kann.

2.6.3.5. Entropie. Für den Carnotschen Kreisprozeß gilt $\overline{Q}_2/Q_1 = T_2/T_1$ oder $Q_1/T_1 + Q_2/T_2 = 0$. Die Summe der „reduzierten Wärmeenergien" verschwindet. Denkt man sich einen beliebigen Kreisprozeß (Abb. 2.6.3–7) in infinitesimale isotherme und adiabatische Zustandsänderungen un-

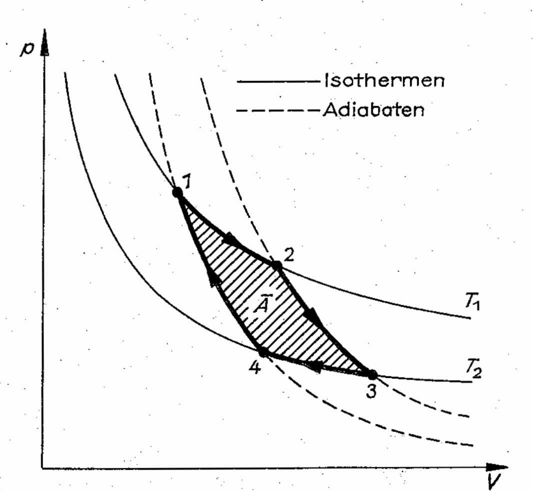


Abb. 2.6.3–4 Carnotscher Kreisprozeß

terteilt, dann gilt bei reversibler Führung $\sum_i Q_i/T_i = 0$. Die reduzierten Wärmeenergien haben eine Eigenschaft, die von den Zustandsgrößen her bekannt ist.
Es kann daher eine neue Zustandsgröße, die Entropie, eingeführt werden. Die *Entropieänderung* bei einer Zustandsänderung ist über

$$\mathrm{d}S = \delta Q_{\mathrm{rev}}/T$$

definiert. Sie kann nur bei reversibler Führung aus der dem System zugeführten Wärmemenge (und der Temperatur) bestimmt werden. Die Entropie ist eine Zustandsgröße, wie aus der oben gefundenen Aussage $\oint \mathrm{d}S = 0$ folgt. Über die

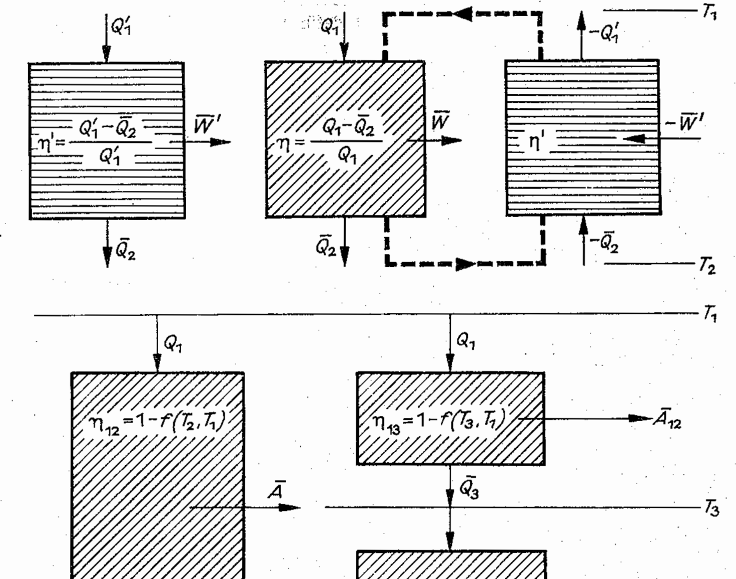


Abb. 2.6.3-5 Kopplung zweier Kreisprozesse (– – –) mit $\eta' > \eta$

Abb. 2.6.3-6 Vergleich eines reversiblen Kreisprozesses zwischen zwei Temperaturen T_1, T_2 mit zwei Prozessen zwischen T_1 und T_3 bzw. T_3 und T_2

Einführung der Entropie ist also eine quantitative Formulierung des zweiten Hauptsatzes gewonnen worden.

2.6.3.6. Formulierungen des zweiten Hauptsatzes. Wie für den ersten Hauptsatz lassen sich auch für den zweiten Hauptsatz der Wärmelehre verschiedene äquivalente Formulierungen angeben: 1. Es ist unmöglich, ein perpetuum mobile zweiter Art zu konstruieren. 2. Der Wirkungsgrad beliebiger reversibler Kreisprozesse zwischen zwei Wärmereservoiren T_1 und T_2 ist gleich groß und gleich dem Wirkungsgrad des Carnotschen Kreisprozesses $\eta = 1 - T_2/T_1$. 3. Die Entropie ist eine Zustandsgröße. Daher ist die Entropieänderung bei einer Zustandsänderung von einem Zustand 1 in einen Zustand 2 unabhängig davon, auf welchem Wege das System von 1 nach 2 geführt wird,

$$\overset{(a)}{\int_1^2} \mathrm{d}S = \overset{(b)}{\int_1^2} \mathrm{d}S = S_2 - S_1.$$

4. Für einen Kreisprozeß gilt $\oint \mathrm{d}S = 0$.

2.6.3.7. Entropie abgeschlossener Systeme. Unter einem *abgeschlossenen System* versteht man ein System, dem weder Arbeit noch Wärme zugeführt wird. Bei der Betrachtung auch anderer Prozesse müßte man hinzufügen, ... dem keine elektrische Energie, keine chemische Energie, keine Masse usw. zu- oder abgeführt wird. Innere Energie und Volumen bleiben also konstant, auch wenn in dem System noch Prozesse ablaufen.
Verlaufen diese Prozesse reversibel, dann kann die Entropieänderung gemäß $\mathrm{d}S = \delta Q_{\mathrm{rev}}/T$ berechnet werden, sie verschwindet wegen $\delta Q = 0$. Bei reversiblen Prozessen ändert sich die Entropie in einem abgeschlossenen System nicht.

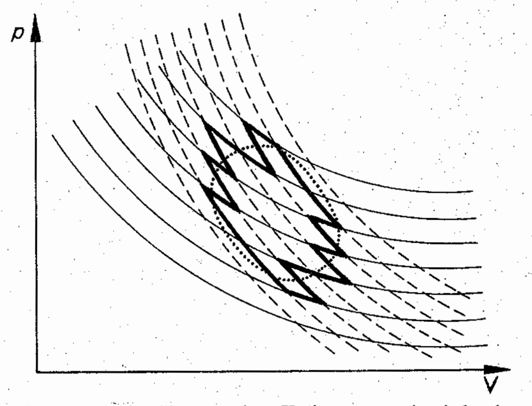


Abb. 2.6.3-7 Annäherung eines Kreisprozesses (·····) durch Isothermen (———) und Adiabaten (– – –)

Bei irreversiblen Prozessen läßt sich aus $\delta Q = 0$ nicht auf $dS = 0$ schließen. Wird der gleiche Zustand auf reversiblem Weg herbeigeführt, dann kann das System eine Arbeit $\delta \bar{W}$ leisten (vgl. 2.6.3.2.), wobei es einem Wärmebad wegen $dU = 0$ die Wärmeenergie $\delta Q = \delta \bar{W} > 0$ entzieht. Die Entropieänderung bei dieser reversiblen Führung ist $dS = \delta Q/T > 0$. Sie ist gleich der Entropieänderung bei irreversibler Führung, da die Entropie eine Zustandsgröße ist und die Endzustände gleich sein sollen. In einem abgeschlossenen System kann die Entropie daher nur zunehmen, nicht abnehmen. Laufen nur reversible Prozesse ab, dann bleibt sie konstant. Es gilt $dS \geqq 0$ im abgeschlossenen System.

Der erste Hauptsatz $dU = \delta Q - p\,dV$ gilt für reversible und irreversible Prozesse. Bei reversiblen Prozessen kann man δQ durch $T dS$ ersetzen, $dU = T dS - p dV$ bzw. $dS = (dU + p dV)/T$. Bei irreversiblen Prozessen entsteht eine größere Zunahme der Entropie, als es durch $\delta Q/T$ angegeben wird. Daher gilt

$$dS - (dU + p\,dV)/T \geqq 0$$

für *beliebige Prozesse* und Systeme. Das Gleichheitszeichen gilt nur bei reversibler Führung.

2.6.4. Thermodynamische Potentiale

Thermische Zustandsgleichung $p(V, T)$ kalorische Zustandsgleichung $U(V, T)$, spezifische Wärme, Kompressibilität usw. sind thermodynamische Eigenschaften eines Systems. Teilweise hängen sie unmittelbar miteinander zusammen und eine Größe kann mit Hilfe der anderen berechnet werden, wie die isotherme Kompressibilität aus der thermischen Zustandsgleichung. Man kann danach fragen, welche Größen man kennen muß, damit das betrachtete System vollständig bestimmt ist und man alle anderen Größen berechnen kann. Es zeigt sich, daß ein System durch eine – geeignet gewählte – Zustandsgröße vollständig charakterisiert werden kann. Diese nennt man ein *thermodynamisches Potential*.

2.6.4.1. Innere Energie als thermodynamisches Potential. Der erste Hauptsatz macht eine Aussage über die Änderung der inneren Energie, $dU = \delta Q - p\,dV$. Der zweite Hauptsatz besagt, daß die Entropie eine Zustandsgröße ist, deren Änderung sich gemäß $dS = \delta Q_{rev}/T$ bestimmt. Werden beide Aussagen zusammengefaßt, $dU = T dS - p dV$, so ergibt sich die Änderung der inneren Energie bei reversibler Prozeßführung aus der Änderung der beiden Zustandsgrößen Entropie und Volumen. Temperatur und Druck treten dabei als Vorfaktoren auf.

Bei zwei Freiheitsgraden des Systems hängen die Zustandsgrößen von zwei Variablen ab. Man kann z.B. V, T oder p, T dafür auswählen, aber auch die Wahl S, V ist möglich. Die Änderung der inneren Energie bei Änderung der Variablen ist mathematisch durch das totale Differential gegeben. Es lautet in den genannten Fällen: $dU = (\partial U/\partial V)_T dV + (\partial U/\partial T)_V dT$ bzw. $dU = (\partial U/\partial p)_T dp + (\partial U/\partial T)_p dT$ bzw. $dU = (\partial U/\partial S)_V dS + (\partial U/\partial V)_S dV$. Die indizierte Variable ist beim Differenzieren jeweils konstant zu halten. Werden diese Ausdrücke mit der aus den Hauptsätzen folgenden Aussage $dU = T dS - p dV$ verglichen, dann liefern in dem Fall, in dem die innere Energie als Funktion von S und V bekannt ist, die partiellen Ableitungen

$$\left(\frac{\partial U(S, V)}{\partial S}\right)_V = T(S, V), \quad -\left(\frac{\partial U(S, V)}{\partial V}\right)_S = p(S, V)$$

unmittelbar die Abhängigkeit der Temperatur und des Druckes von diesen Variablen. $T = T(S, V)$ kann nach der Entropie aufgelöst werden, $S = S(V, T)$. Damit kann man die Entropie in $U(S, V)$ und in $p(S, V)$ eliminieren, womit die Abhängigkeit der inneren Energie $U(V, T)$ bzw. des Druckes $p(V, T)$ von den Variablen V, T vorliegt. In analoger Weise kann man alle interessierenden Größen in Abhängigkeit von einem beliebig wählbaren Variablenpaar bestimmen.

Die innere Energie, gegeben als Funktion der Entropie und des Volumens $U(S, V)$ ist ein thermodynamisches Potential. Aus dieser Funktion können alle Eigenschaften des Systems berechnet werden. Für das ideale Gas lautet diese Funktion

$$U(S, V) = U_0 \exp[(S/z - \sigma - R \ln V/V_0)/C_V]$$

mit der Konstanten $\sigma = S(U_0, V_0)/z$. Durch Differentiation und Umrechnen auf die Variablen V und T folgen daraus $U = zC_V T$ und $pV = zRT$.

2.6.4.2. Wechsel der Variablen. Je nach der Prozeßführung wählt man verschiedene Zustandsgrößen als Variable, um die Zustandsänderungen einfach beschreiben und untersuchen zu können. Bei isobaren Prozessen wählt man den Druck als eine der beiden Variablen, bei isothermen Prozessen die Temperatur usw. Dann kann man die bei dem Prozeß auftretenden Veränderungen durch die Variation der zweiten Variablen allein beschreiben.

Das thermodynamische Potential $U(S, V)$ ist nur dann ein günstiger Ausgangspunkt, wenn man Entropie und Volumen als Variable wählen will. Das ist meist nicht der Fall. Es lassen sich aber leicht andere Funktionen definieren, die in Abhängigkeit von den gewünschten Variablen thermodynamische Potentiale sind. Die wichtigsten sind *freie Energie* $F(V, T)$, *Enthalpie* $H(S, p)$ und *freie Enthalpie* $G(p, T)$. Deren Ableitungen führen auch unmittelbar auf Zustandsgrößen, wie aus Tab. 2.6.4–1 zu ersehen ist.

Für das *van-der-Waals-Gas* ist die freie Energie

$$F(T, V) = zC_V T(1 - \ln T/T_0 - \sigma/C_V) - zRT \ln[(V - zb)/(V_0 - zb)] - z^2a^2/V.$$

Dabei wurde der Einfachheit halber C_V als temperaturunabhängig betrachtet. Aus dem thermodynamischen Potential $F(T, V)$ folgen z.B. $U = zC_V T - z^2a^2/V$ und $p = zRT/(V - zb) - z^2a^2/V^2$.

Tab. 2.6.4–1 Thermodynamische Potentiale

Bezeichnung	Symbol und Definition	Zustandsänderung	Ableitung	
Entropie	$S(U, V)$	$dS = (dU + p dV)/T$	$(\partial S/\partial U)_V = 1/T$,	$(\partial S/\partial V)_U = p/T$
innere Energie	$U(S, V)$	$dU = T dS - p dV$	$(\partial U/\partial S)_V = T$,	$(\partial U/\partial V)_S = -p$
freie Energie	$F(T, V) = U - TS$	$dF = -S dT - p dV$	$(\partial F/\partial T)_V = -S$,	$(\partial F/\partial V)_T = -p$
Enthalpie	$H(S, p) = U + pV$	$dH = T dS + V dp$	$(\partial H/\partial S)_p = T$,	$(\partial H/\partial p)_S = V$
freie Enthalpie	$G(p, T) = F + pV$	$dG = -S dT + V dp$	$(\partial G/\partial T)_p = -S$,	$(\partial G/\partial p)_T = V$

2.6.4.3. Thermodynamische Ungleichungen. Der zweite Hauptsatz der Wärmelehre sagt aus, daß die Entropie in einem abgeschlossenen System nur zunehmen oder konstant bleiben kann. Im thermodynamischen Gleichgewicht hat sie ihr Maximum erreicht, es finden keine irreversiblen (Ausgleichs-)Prozesse mehr statt. Die Bedingung „abgeschlossenes System" bedeutet, daß der Energieinhalt U und das Volumen V fest vorgegeben sind, sich im Zeitablauf nicht ändern. Die Bedingung, daß die Entropie ein Maximum annimmt, liefert die Art des Gleichgewichtszustandes. Werden Druck und Temperatur oder Volumen und Temperatur usw. vorgegeben, dann lassen sich die Bedingungen für den Gleichgewichtszustand mit Hilfe der thermodynamischen Potentiale leicht formulieren (Tab. 2.6.4–2). Wenn z. B. mehrere Komponenten unter gegebenen Druck in einem Schmelztiegel erhitzt werden, so wird die Art der Phasen, die Konzentration der einzelnen Komponenten in den Phasen usw. im thermodynamischen Gleichgewichtszustand durch die Forderung bestimmt, daß die freie Enthalpie ein Minimum annehmen muß.

Tab. 2.6.4–2 Bedingungen für den thermodynamischen Gleichgewichtszustand

Nebenbedingung	Gleichgewichtsbedingung
abgeschlossenes System	$S \rightarrow$ Maximum
T = konst., V = konst.	$F \rightarrow$ Minimum
T = konst., p = konst.	$G \rightarrow$ Minimum

2.6.5. Thermodynamische Prozesse

Ideales Gas, van-der-Waalssches Gas und isotroper Festkörper sind wichtige Modellsysteme, an denen thermodynamische Prozesse untersucht werden können.

2.6.5.1. Volumenabhängigkeit der inneren Energie eines Gases. Beim Gay-Lussac-Versuch (Abb. 2.6.5–1) befindet sich im Ausgangszustand in einer Flasche (I) ein Gas unter hohem Druck im thermodynamischen Gleichgewicht. Die zweite Flasche ist leer. Das Kalorimeter (zur Messung von Wärmeenergien) und das Gas haben die Temperatur T. Nach Öffnen des Ventils findet ein Druckausgleich statt, ohne daß Arbeit abgegeben wird, $\Delta W = 0$. Im thermodynamischen Gleichgewicht befindet sich das Gas bei konstantem Druck und konstanter Temperatur in einem doppelt so großen Volumen.

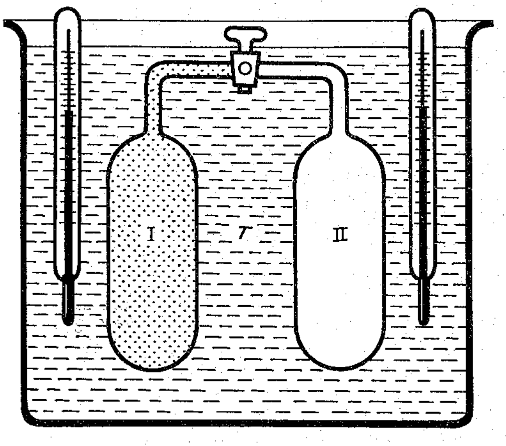


Abb. 2.6.5–1 Gay-Lussac-Versuch

Je nach der Volumenabhängigkeit der inneren Energie $U(T, V)$ wird nach dem ersten Hauptsatz $\Delta U = U(T', 2V) - U(T, V) = \Delta Q + \Delta W = \Delta Q$ dem Kalorimeter Wärme entzogen bzw. Wärmeenergie an das Kalorimeter abgegeben. Diese läßt sich aus der Änderung der Temperatur $T \rightarrow T'$ des Kalorimeters bestimmen. Man beobachtete im Rahmen der Meßgenauigkeit keine Temperaturänderung $T = T'$, d. h. $\Delta Q = 0$. Daher ist die innere Energie des Gases unabhängig vom Volumen. Genauere Messungen zeigten, daß diese Aussage nur in der Näherung richtig ist, in der man

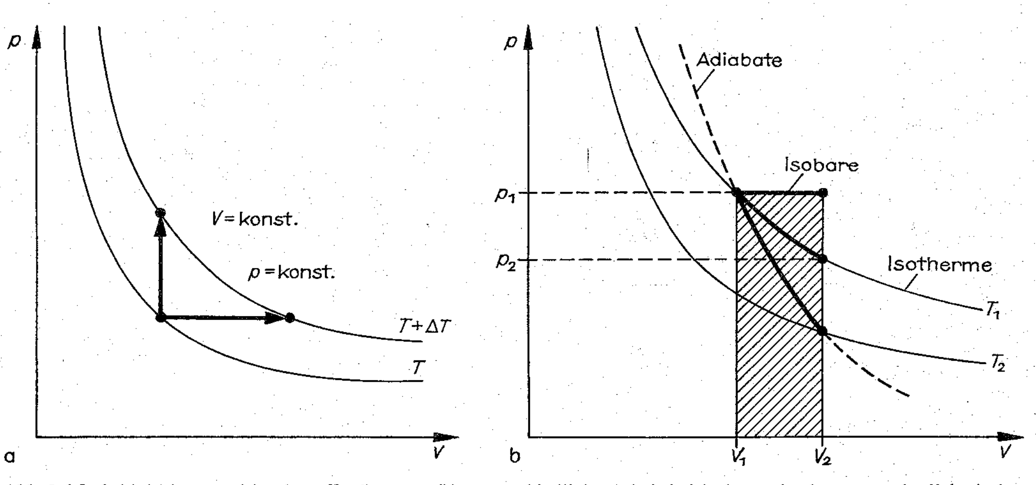


Abb. 2.6.5–2 (a) isobare und isochore Erwärmung, (b) unterschiedliche Arbeit bei isobarer, isothermer und adiabatischer Expansion

das Gas als ideales Gas ansehen kann. Beim van-der-Waals-Gas hängt die innere Energie vom Volumen ab. Die *innere Energie des idealen Gases* ist unabhängig vom Volumen, sie ist nur eine Funktion der Temperatur

$U(T) = zC_V T.$

Aus dem zweiten Hauptsatz folgt ein Zusammenhang zwischen der Volumenabhängigkeit der inneren Energie und der Zustandsgleichung

$(\partial U/\partial V)_T = T^2[\partial(p/T)/\partial T]_V.$

Für das ideale Gas ergibt sich daraus $\partial U/\partial V)_T = 0$, da $p/T = zR/V$ bei festgehaltenem Volumen unabhängig von der Temperatur ist. Beim *van-der-Waals-Gas* ist die innere Energie $U = zC_V T - z^2a^2/V$, woraus $\partial U/\partial V = z^2a^2/V^2$ folgt. Das stimmt mit dem aus $p = zRT/(V - zb) - z^2a^2/V^2$ folgenden Wert für $T^2\partial(p/T)/\partial T$ überein.

2.6.5.2. Spezifische Wärme. Fragt man nach der Wärmeenergie ΔQ, die benötigt wird, um eine bestimmte Temperaturerhöhung ΔT zu erzielen, dann muß man bei einem System mit zwei Freiheitsgraden festlegen, unter welcher Bedingung die Temperatur erhöht wird. So kann man z. B. die Temperatur (Abb. 2.6.5–2a) bei konstantem Druck oder konstantem Volumen vergrößern. Den Quotienten aus zugeführter Wärmeenergie und Temperaturerhöhung bezeichnet man als *Wärmekapazität*. Um eine von der Menge der Substanz unabhängige Materialeigenschaft zu haben, bezieht man die Wärmekapazität auf ein Mol [*Molwärme C*, Einheit J/(mol · K)] bzw. führt die *spezifische Wärme c* (Einheit J/kg · K) ein (Tab. 2.6.5–3). Der Meßvorschrift entsprechend gilt

$C_V = (\Delta Q/\Delta T)_V = \text{konst.}, \quad C_p = (\Delta Q/\Delta T)_p = \text{konst.}$

Tab. 2.6.5–3 Molwärmen bei 25 °C, Angaben in J/(mol · K) (nach LANDOLT-BÖRNSTEIN)

Substanz	C_p	C_V	Substanz	C_p
Ag	25,49	24,27	Edelgase	20,79
Al	24,34	23,03	H_2 (25 K)	20,79
Au	25,38	24,47	H_2	28,83
Cu	24,47	23,63	O_2	29,36
Li	23,64	22,53	O_3	39,20
Pb	26,82	24,84	N_2	29,08
Pt	26,57	24,95	NH_3	35,52
W	24,08	23,59	CO_2	37,13

Spezifische Wärme kann man mit dem *Kalorimeter* messen. In einem wärmeisolierten Gefäß (vgl. Abb. 2.6.5–1) befindet sich eine bekannte Menge einer Flüssigkeit. Dieser wird eine definierte Energiemenge zugeführt, z. B. die in einem elektrischen Widerstand innerhalb der Zeit Δt erzeugte Joulesche Wärme $I^2R\Delta t$. Die dadurch bewirkte Temperaturerhöhung kann abgelesen werden. Bei sorgfältiger Durchführung des Versuches muß beachtet werden, daß auch das Gefäß und der Rührer (durch Rühren wird die zugeführte Wärmemenge rascher auf die gesamte Flüssigkeit verteilt) Wärmeenergie aufnehmen und daß die Isolation unvollkommen ist.

Kennt man die Wärmekapazität des Kalorimeters einschließlich Flüssigkeit, dann kann man durch Eintauchen eines Probekörpers bekannter Masse und Temperatur dessen spezifische Wärme aus der sich einstellenden Temperatur T_m gewinnen. Eine Flüssigkeit, deren spezifische Wärme bestimmt werden soll, darf nicht einfach mit der Kalorimeterflüssigkeit gemischt werden, da die Mischungswärme das Ergebnis verfälschen kann. Man denke z. B. an die Erwärmung, die auftritt, wenn man konzentrierte Schwefelsäure in Wasser gießt.

Aus den Hauptsätzen der Wärmelehre folgt, daß die Molwärmen $C_V = (\partial U/\partial T)_V$, $C_p = (\partial H/\partial T)_p$ in einfacher Weise aus der inneren Energie bzw. der Enthalpie berechnet werden können. Die Differenz der Molwärmen

$C_p - C_V = T(\partial p/\partial T)_V(\partial V/\partial T)_p,$

$C_p - C_V = R$ für das ideale Gas,

ist aus der thermischen Zustandsgleichung bestimmbar. Aus gaskinetischen Betrachtungen (vgl. 2.6.12.) folgt für ein einatomiges Gas $C_V = 3R/2$ und $C_p = 5R/2 = 20{,}79$ J/(mol · K) und für zweiatomige Gase $C_V = 5R/2$ bzw. $C_p = 7R/2 = 29{,}10$ J/(mol · K). Der häufig auftauchende Quotient C_p/C_V, auch *Adiabatenexponent* γ genannt, hat daher den Wert $\gamma_{\text{einat.}} = 5/3$, $\gamma_{\text{zweiat.}} = 1{,}4$. Die Abweichungen von C_p z. B. für H_2, O_2 und N_2 vom Wert $7R/2$ können erklärt werden, wenn man Rotations- und Schwingungsbewegungen quantenphysikalisch erfaßt. Der Unterschied zwischen C_p und C_V rührt daher, daß die Erwärmung bei konstantem Druck mit einer Volumenausdehnung verknüpft ist. Das System leistet daher Arbeit. Diese muß von der zugeführten Energiemenge und von der inneren Energie aufgebracht werden. Ist die innere Energie vom Volumen unabhängig, wie das beim idealen Gas der Fall ist, dann geht die Ausdehnungsarbeit voll zu Lasten der zugeführten Wärmeenergie. Bei Festkörpern und Flüssigkeiten ist wegen der geringen Wärmeausdehnung der Unterschied zwischen C_p und C_V oft vernachlässigbar (vgl. Tab. 2.6.5–3). Die Volumenvergrößerung beim Schmelzen beträgt wenige Prozent, z. B. 5,4 % bei Silber, 6,9 % bei Aluminium und 4,4 % bei Kupfer.

Bei einer isochoren Erwärmung ist $\delta Q = C_V \mathrm{d}T$, bei einer isobaren Erwärmung $\delta Q = C_p \mathrm{d}T$ und bei einem adiabatischen Prozeß gilt definitionsgemäß $\delta Q = 0$. Bei einer *isothermen Expansion* (vgl. Abb. 2.6.5–2b), folgt aus den Hauptsätzen, daß das System die Wärmeenergie

$\delta Q = T(\partial p/\partial T)_V \mathrm{d}V, \quad \Delta Q_{\text{ideales Gas}} = zRT \ln V_2/V_1$

aufnimmt. Für das ideale Gas wurde dieser Ausdruck integriert.

2.6.5.3. Wärmeausdehnung, Kompressibilität. Zu jeder Substanz gehört eine Zustandsgleichung der Form $V = V(p, T)$, falls die Substanz homogen und isotrop ist. Dieser Zusammenhang zwischen Volumen, Druck und Temperatur spiegelt die Beobachtung wider, daß sich ein Körper beim Erwärmen ausdehnt, bei Erhöhung des Druckes komprimiert wird usw. Da das betrachtete System zwei Freiheitsgrade hat, muß immer genau angegeben werden, wie die Zustandsänderung erfolgen soll.

Der *Ausdehnungskoeffizient* beschreibt die relative Volumenänderung mit der Temperatur bei konstantem Druck $\alpha = (1/V)\,(\partial V/\partial T)_p$.

Die *Kompressibilität* drückt die Volumenänderung bei Erhöhen des Druckes aus. Zwei Fälle treten häufig auf:

1. Die Druckerhöhung erfolgt bei konstanter Temperatur, *isotherme Kompressibilität* $\varkappa_T = (1/V)\,(\partial V/\partial p)_T$.
2. Während der Kompression kann kein Wärmeaustausch stattfinden, z. B. weil diese zu schnell erfolgt. Der Prozeß verläuft also adiabatisch, *adiabatische Kompressibilität* $\varkappa_S = -(1/V)\,(\partial V/\partial p)_S$.

Den reziproken Wert der Kompressibilität bezeichnet man als *Kompressionsmodul*. Der *Spannungskoeffizient* be-

schreibt die Druckänderung bei Erhöhung der Temperatur bei festem Volumen $(1/p)\,(\partial p/\partial T)_V = (1/p)\,(\alpha/\varkappa_T)$.

Bei der *Wärmeausdehnung von Gasen* mißt man in guter Näherung (soweit man das Gas als ideal betrachten kann) $\alpha(\vartheta = 0) = 1/273{,}15\,\mathrm{K} = 0{,}003\,661\,\mathrm{K}^{-1}$ und ein lineares Anwachsen des Volumens mit der Temperatur $V(\vartheta, p = \text{konst.}) = V(0\,°\mathrm{C})\,(1 + \alpha\vartheta)$, wobei der Ausdehnungskoeffizient α für alle Gase den gleichen Wert aufweist. Dieses Ergebnis führt mit zur Zustandsgleichung des idealen Gases.

Flüssigkeiten haben einen geringeren Ausdehnungskoeffizienten als Gase. Auch muß seine Temperaturabhängigkeit beachtet werden. Für Quecksilber hat er bei Zimmertemperatur den Wert $0{,}000\,18\,\mathrm{K}^{-1}$. Wasser verringert im Bereich zwischen 0 und 4 °C beim Erwärmen sein Volumen, erst oberhalb von 4 °C wird ein Anstieg des Volumens registriert. Es hat daher auch bei 4 °C seine größte Dichte. Das führt dazu, daß ein Teich an der Oberfläche zuzufrieren beginnt, während sich am Boden Wasser von 4 °C befindet.

Bei *Festkörpern* gibt man den linearen Ausdehnungskoeffizienten α_1 (Tab. 2.6.5–4), d. h. die Änderung von Länge, Breite oder Höhe, an $\alpha_1 = (1/l)\,(\Delta l/\Delta T)$. Der räumliche Ausdehnungskoeffizient $\alpha = 3\alpha_1$ hat bei isotropen Körpern den dreifachen Wert. Auch hier muß man die Temperaturabhängigkeit beachten. Zwischen 0 K und dem Schmelzpunkt vergrößern die meisten Festkörper ihr Volumen um etwa 7 %.

Tab. 2.6.5–4 Linearer Ausdehnungskoeffizient und isotherme Kompressibilität bei 20 °C und Normaldruck, Angaben in $10^{-6}\,\mathrm{K}^{-1}$ bzw. $10^{-11}\,\mathrm{Pa}^{-1}$ (nach SCHULZE und KITTEL)

Substanz	α_1	$\varkappa_T$
Ag	18,7	0,993
Al	23,8	1,385
Au	14,2	0,577
C	0,6...4,3	
Cu	16,8	0,73
Fe	12,1	0,594
Li	56	8,62
Ni	13,1	0,538
Pb	29,2	2,33
Pt	21,4	0,359
Si	2,33	1,012
W	19,3	0,309

Die unterschiedliche Ausdehnung verschiedener Metalle macht man sich beim *Bimetallstreifen* zunutze. Werden z. B. zwei Streifen aus Invar bzw. Messing bei 20 °C spannungsfrei verbunden, dann verbiegt sich der Bimetallstreifen bei höheren Temperaturen wegen des größeren Ausdehnungskoeffizienten von Messing nach der Invarseite hin, bei Temperaturen unter 20 °C nach der Messingseite hin. Invar ist ein Sonderwerkstoff mit einem kleinen Ausdehnungskoeffizienten, der dort eingesetzt werden kann, wo die Wärmeausdehnung unerwünscht ist. Bei *Verbundwerkstoffen* (z. B. Eisen in Beton) oder beim Verschmelzen von Werkstoffen (z. B. Metalldrähte im Glassockel einer Glühbirne) muß man auf gleiche Ausdehnungskoeffizienten der Komponenten achten. Sonst treten bei Temperaturänderungen wegen der unterschiedlichen Ausdehnung große Spannungen auf, die zur Zerstörung führen können. *Wärmespannungen* treten auch in einem homogenen System auf, wenn dessen Teile auf unterschiedliche Temperatur gebracht werden. Feuerfestes Glas zeichnet sich u. a. durch einen kleinen Ausdehnungskoeffizienten aus, wodurch solche Wärmespannungen klein gehalten werden.

Entsprechend der Zustandsgleichung des idealen Gases $pV = zRT$ reduziert sich das Volumen bei Verdopplung des Druckes und bei konstanter Temperatur auf die Hälfte. Die isotherme Kompressibilität ergibt sich aus der Zustandsgleichung zu $\varkappa_T = 1/p$. Bei einer adiabatischen Kompression gilt die Adiabatengleichung $pV^\gamma = \text{konst.}$ Daraus ergibt sich für ideale Gase $\varkappa_S/\varkappa_T = \gamma$. Flüssigkeiten und Festkörper setzen einer Kompression einen wesentlich größeren Widerstand entgegen, die Kompressibilität ist klein (vgl. Tab. 2.6.5–4). Die adiabatische Kompressibilität kann man entsprechend $\varkappa_S = \varkappa_T - T\alpha^2/(\varrho c_p)$ berechnen.

2.6.5.4. Adiabatengleichung. Beim adiabatischen Prozeß ist das System definitionsgemäß wärmeisoliert, $\delta Q = 0$. Die innere Energie kann sich nur dadurch ändern, daß dem System Arbeit zugeführt wird bzw. daß es Arbeit leistet, $\mathrm{d}U + p\,\mathrm{d}V = 0$. Unter Verwendung der aus dem zweiten Hauptsatz folgenden Aussagen über die Volumenabhängigkeit der inneren Energie und die Differenz $C_p - C_V$ folgt $\mathrm{d}V = -\mathrm{d}T(\partial V/\partial T)_p/(\gamma - 1)$ mit $\gamma = C_p/C_V$ für den Zusammenhang zwischen Volumen- und Temperaturänderung beim adiabatischen Prozeß. Die Auswertung liefert für das ideale Gas

$TV^{\gamma-1} = \text{konst.}$ oder $pV^\gamma = \text{konst.}$

Aus dieser Adiabatengleichung wird die Bezeichnung *Adiabatenexponent* für γ (im Falle des idealen Gases) verständlich. Die Adiabaten verlaufen stets steiler als die Isothermen. Wegen $\gamma - 1 > 0$ ist aus $TV^{\gamma-1} = \text{konst.}$ ersichtlich, daß sich das System mit wachsendem Volumen abkühlt. Das wird auch daraus verständlich, daß die Ausdehnungsarbeit zu Lasten der inneren Energie geht.

2.6.5.5. Ausdehnungsarbeit beim idealen Gas. Bei einer Volumenänderung von V auf $V + \mathrm{d}V$ ist die dem Gas zugeführte Arbeit $\delta W = -p\,\mathrm{d}V$, die vom Gas abgegebene Arbeit $\delta\overline{W} = p\,\mathrm{d}V$. Im p-V-Diagramm ist $\overline{W}$ gleich der Fläche unter der Kurve, die die Zustandsänderung beschreibt. In Abb. 2.6.5–2b sind dies die Kurven $p = \text{konst.}$ (Isobaren), $pV = \text{konst.}$ (Isothermen) bzw. $pV^\gamma = \text{konst.}$ (Adiabaten). Der Flächeninhalt hat die Werte:

isobar: $\overline{W} = p_1(V_2 - V_1)$,
isotherm: $\overline{W} = zRT\ln(V_2/V_1) = zRT\ln(p_1/p_2)$,
adiabatisch: $\overline{W} = zR(T_1 - T_2)/(\gamma - 1)$.

Bemerkenswert ist, daß die bei der adiabatischen Expansion abgegebene Arbeit nur von Anfangs- und Endtemperatur abhängt.

2.6.5.6. Carnotscher Kreisprozeß mit idealem Gas. Ist beim Carnotschen Kreisprozeß der Arbeitsstoff ein ideales Gas, dann sind die auf den Teilabschnitten (vgl. Abb. 2.6.3–4) zu- bzw. abgeführten Wärmeenergien und die Expansions- bzw. Kompressionsarbeit gegeben durch:

$$Q_{12} = zRT_1\ln(V_2/V_1), \quad \overline{W}_{12} = zRT_1\ln(V_2/V_1),$$
$$Q_{23} = 0, \quad \overline{W}_{23} = zR(T_1 - T_2)/(\gamma - 1),$$
$$\overline{Q}_{34} = zRT_2\ln(V_3/V_4), \quad W_{34} = zRT_2\ln(V_3/V_4),$$
$$Q_{41} = 0. \quad W_{41} = zR(T_1 - T_2)/(\gamma - 1).$$

Weil die innere Energie eine Zustandsgröße ist, ist die Summe aller dem System bei einem Kreisprozeß zugeführten Arbeiten und Wärmeenergien gleich Null. Da die Punkte 1 und 4 bzw. 2 und 3 auf der gleichen Adiabaten liegen, gilt $T_1V_1^\gamma = T_2V_4^\gamma$ bzw. $T_1V_2^\gamma = T_2V_3^\gamma$. Daraus folgt die

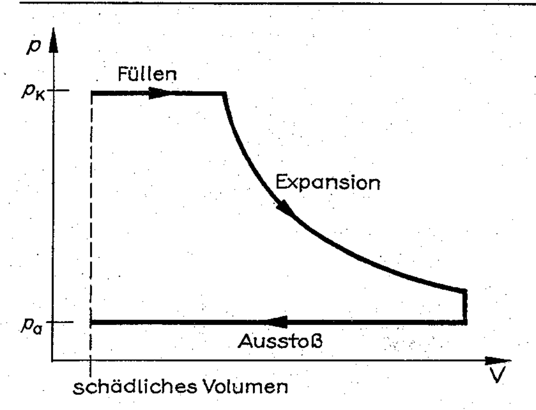


Abb. 2.6.5–5 Schematisches Indikatordiagramm einer Dampfmaschine; p_K = Kesseldruck, p_a = äußerer Druck

Gleichheit der Volumenverhältnisse $V_2/V_1 = V_3/V_4$. Im Wirkungsgrad $\eta = \overline{W}/Q_{12} = (T_1 - T_2)/T_1$ kürzen sich alle Faktoren neben den Temperaturen heraus, η hängt nur von T_1 und T_2 ab.

2.6.5.7. Wärmekraftmaschinen. Das in heißen, unter hohen Druck stehenden Gasen enthaltene Vermögen, Arbeit zu leisten, kann auf verschiedene Weise genutzt werden. In den Kolbendampfmaschinen wird die Arbeit durch den Druck auf den sich verschiebenden Kolben abgegeben, in den Gasturbinen wird die kinetische Energie des strömenden Gases ausgenutzt und bei den Strahltriebwerken führt der Rückstoß zur Beschleunigung.

Kolbendampfmaschinen benötigen einen Dampfkessel, in dem Wasser zum Sieden gebracht wird, so daß Heißdampf unter hohem Druck entsteht. Typische Werte sind 600 °C und 20 MPa. Über eine Ventilsteuerung, die möglichst kleine schädliche Räume aufweisen soll, wird Dampf in den Zylinder eingelassen. Bei geschlossenen Ventilen expandiert dieser unter Abgabe von Arbeit, bis der Dampf nahezu den Außendruck erreicht hat. Dann wird der Zylinder über ein Auspuffventil entleert. Abb. 2.6.5–5 zeigt schematisch das Arbeitsdiagramm. Während des Füllens verschiebt sich der Kolben bei konstantem Druck. Nach Schließen des Ventils nimmt der Druck bei der Expansion ab. Ohne die Wärmeverluste über die Zylinderwände wäre dies eine adiabatische Expansion. Das Leeren des Zylinders erfordert einen Arbeitsaufwand der Dampfmaschine. Ein hoher Wirkungsgrad erfordert eine hohe Anfangstemperatur T_1. Sie wird durch die Druckfestigkeit des Dampfkessels begrenzt. Weiterhin soll T_2 möglichst niedrig liegen, d.h. der Außendruck möglichst klein sein. Wird der Dampf in die Außenluft abgestoßen, so kann T_2 nicht unter der Sättigungstemperatur des Wasserdampfes bei Normaldruck (≈0,1 MPa) liegen. Arbeitet man mit einem Kondensator, dann ist dies die Kühlwassertemperatur. Man kann dadurch die Entspannung bis 10 kPa führen. Kolbendampfmaschinen erreichen einen Wirkungsgrad von 10 bis 15 %, d.h. nur etwa 10 % der beim Feuern frei werdenden Wärmeenergie werden als Arbeit genutzt.

Gasturbinen wandeln einen Teil der Energie des hochgespannten Dampfes in kinetische Energie strömenden Gases um. In der Düse findet eine genähert adiabatische Expansion statt. Bei einer sich verjüngenden Düse ergibt sich die Schallgeschwindigkeit als Grenzwert für die Geschwindigkeit des stationär ausströmenden Gases. Bei der Laval-Düse, bei der sich an den engsten Querschnitt noch eine Erweiterung anschließt, kann der Dampf jedoch mit Über-

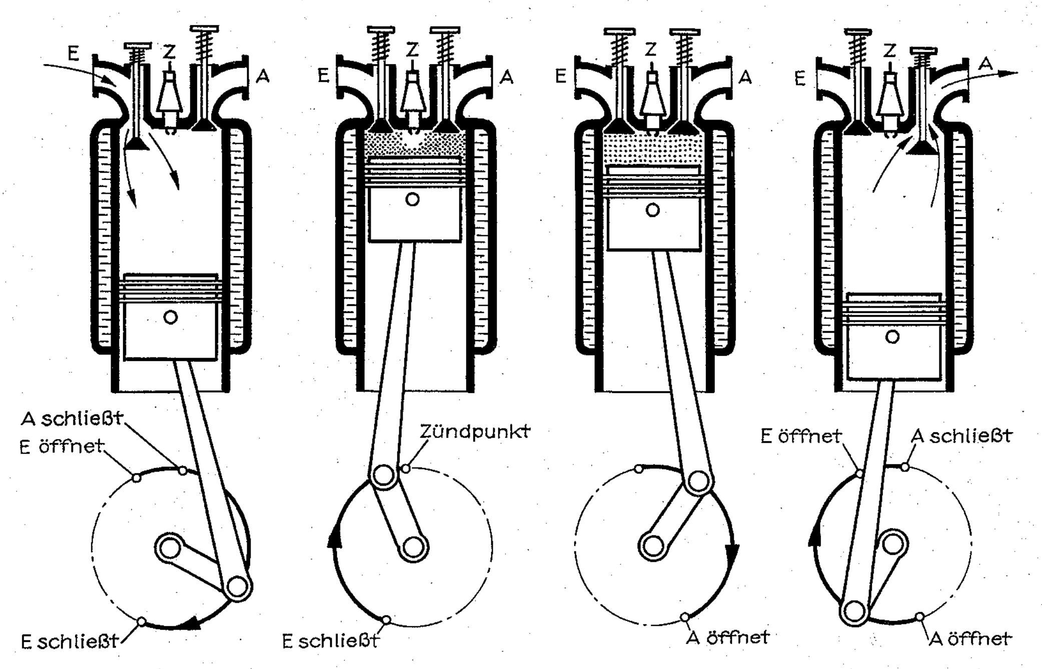


Abb. 2.6.5–6 Viertakt-Ottomotor, E und A Ventile, Z Zündung

schallgeschwindigkeit ausströmen. Der Dampf trifft auf die Schaufeln der Laufräder. Bei geeigneter Form der Schaufeln und einer Laufgeschwindigkeit, die etwa halb so groß wie die Strahlgeschwindigkeit ist, verläßt der Dampf das Laufrad mit einer Geschwindigkeit nahe bei Null. In zusammengesetzten Turbinenausführungen wird die Geschwindigkeit stufenweise abgebaut.
Wenn man von den Verlusten absehen könnte, würde sich der Wirkungsgrad der Dampfturbine nicht von dem der Kolbendampfmaschine unterscheiden. Bei großer Gesamtleistung ist die Turbine der Kolbendampfmaschine überlegen. Der Wirkungsgrad kann 30 % erreichen.
Verbrennungsmotoren erzeugen den hochgespannten Heißdampf nicht in einem Dampfkessel, sondern durch Verbrennen eines Luft-Kraftstoff-Gemisches im Zylinder selbst. Abb. 2.6.5–6 zeigt die Arbeitsweise des *Otto-Viertaktmotors*. Im ersten Takt ist das Ventil zum Vergaser hin geöffnet, das Kraftstoff-Luft-Gemisch wird angesaugt. Im zweiten Takt wird dieses Gemisch verdichtet. Kurz vor Erreichen der höchsten Verdichtung wird es gezündet. Druck und Temperatur erhöhen sich sprunghaft. Der dritte Takt ist der Arbeitstakt. Das verbrannte heiße Gas (1 000 – 2 000 °C) dehnt sich aus und kühlt sich ab. Im vierten Takt werden die verbrannten Gase ausgestoßen. Abb. 2.6.5–7 zeigt das Indikatordiagramm und die Kraftübertragung auf die Kurbel. Die Expansion wird nicht bis zu dem minimal möglichen Druck geführt, da im letzten Teil (schraffiert) nur noch wenig Arbeit gewonnen wird. Damit der Motor gleichmäßig läuft, werden zwei, vier oder mehr Zylinder mit zeitlich versetzter Zündung eingebaut. Beim *Otto-Zweitaktmotor* (Abb. 2.6.5–8) wird in jedem zweiten Takt Arbeit geleistet. Am Ende des Arbeitstaktes strömt das verbrannte Gas wegen seines noch relativ hohen Druckes durch den Auspuffschlitz ab. Gleichzeitig wird der Zylinder

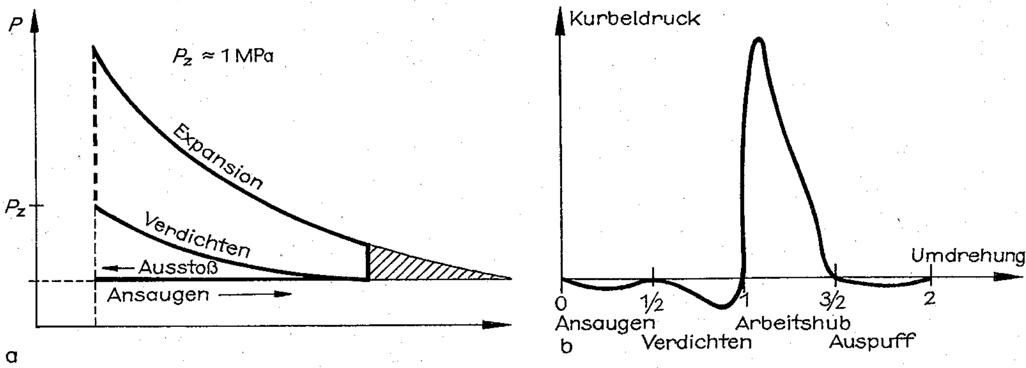


Abb. 2.6.5–7 (a) schematisches Indikatordiagramm und (b) Kraftübertragung beim Otto-Viertaktmotor

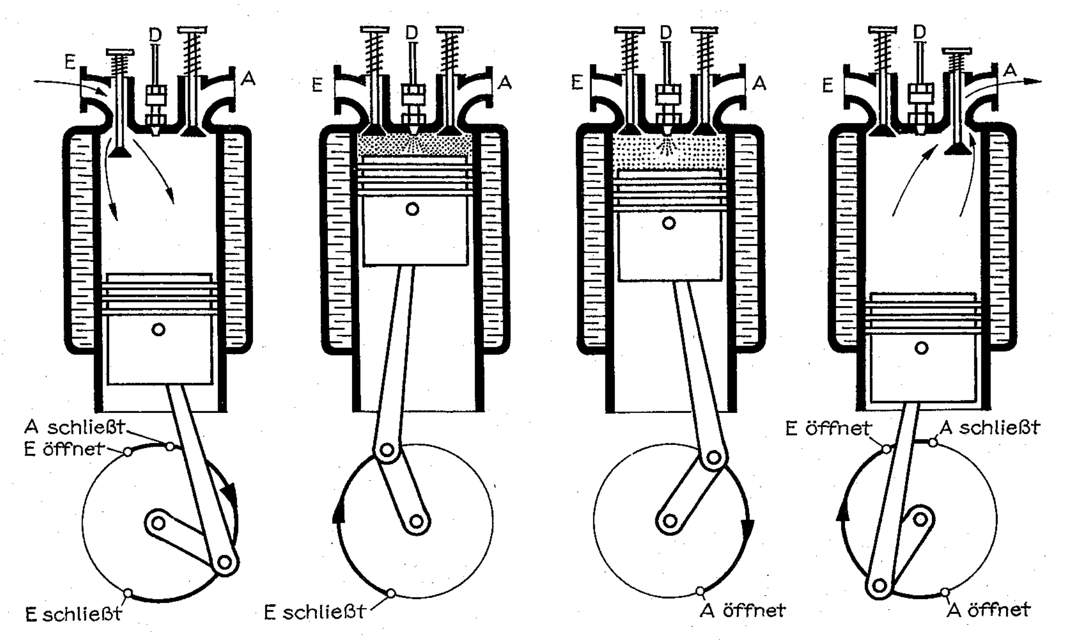


Abb. 2.6.5–9 Diesel-Motor

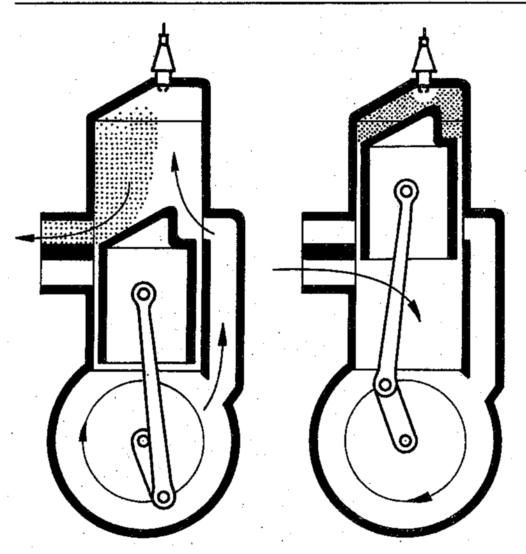


Abb. 2.6.5–8 Otto-Zweitaktmotor

durch einströmendes Kraftstoffgemisch gespült. Danach tritt die Verdichtung ein. Nach der Zündung – kurz vor Erreichen der höchsten Verdichtung – folgt wieder ein Arbeitstakt.

Der Wirkungsgrad des Otto-Motors wächst mit dem Wert der Verdichtung vor der Zündung und mit der Temperatur des verbrannten Kraftstoffgemisches nach der Zündung. Der Verdichtung (bis etwa 1 MPa) sind durch die Selbstentzündung Grenzen gesetzt. Die um 1 000 °C liegende Temperatur des verpufften Kraftstoffgemisches ist durch das Luft-Kraftstoff-Verhältnis beeinflußbar. Eine richtige Vergasereinstellung hilft also, Kraftstoff zu sparen.

Beim *Diesel-Motor* (Abb. 2.6.5–9) wird die im ersten Takt angesaugte Luft ohne Kraftstoff verdichtet. Da bei reiner Luft keine Explosionsgefahr besteht, wird bis auf etwa 4 MPa verdichtet. In die heiße hochgespannte Luft wird das Dieselöl unter hohem Druck (20 bis 30 MPa) in Form kleiner Flüssigkeitströpfchen eingesprüht. Es verbrennt. In dieser Anfangsphase des dritten Taktes bleibt der hohe Druck

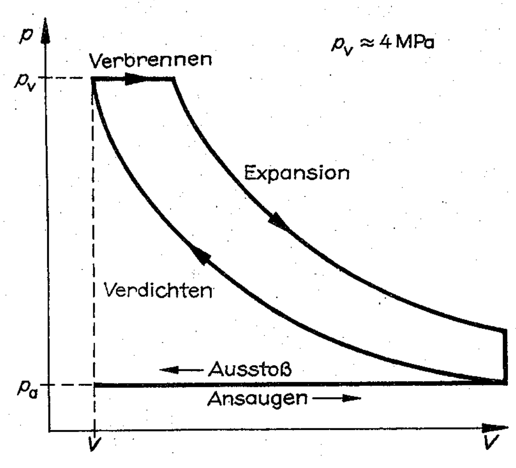


Abb. 2.6.5–10 Schematisches Indikatordiagramm des Diesel-Motors

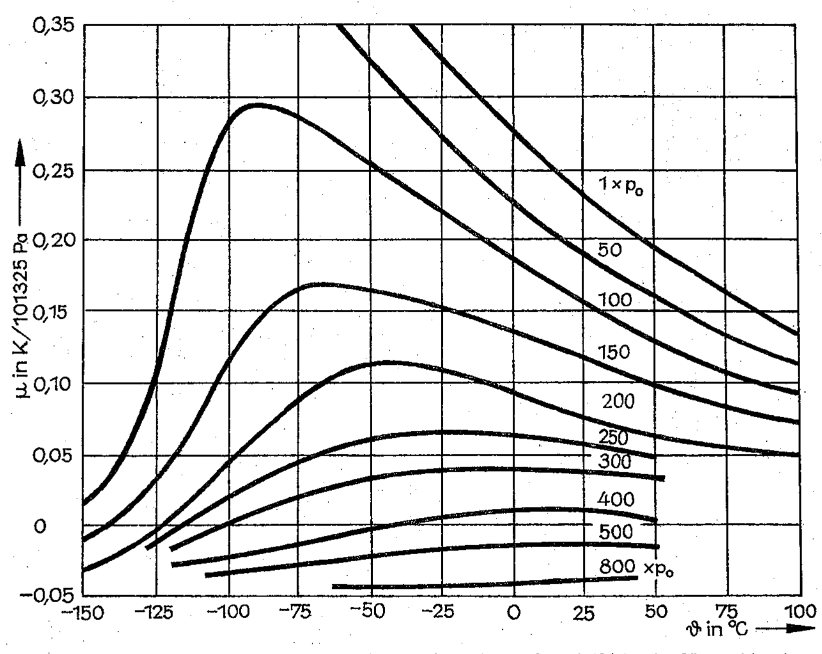


Abb. 2.6.5–12 Joule-Thomson-Koeffizient von Luft, Isobaren für Vielfache des Normaldruckes $p_0 = 101\,325$ Pa

aufrecht erhalten (Abb. 2.6.5–10). Danach erfolgt eine mit Temperatur- und Druckerniedrigung verbundene Expansion. Im vierten Takt wird das verbrannte Gas ausgestoßen. Im Diesel-Motor können schwerverdampfende, billige Brennstoffe eingesetzt werden. Wegen des hohen Anfangsdruckes erreicht der Wirkungsgrad Werte über 40 %.

Strahltriebwerke von Flugzeugen nutzen den Rückstoß ausströmender Gase (verbrannter Kraftstoff und angesaugte Luft) zur Beschleunigung. Eine Turbine wird nur benötigt, um den Verdichter zu betreiben.

Raketen arbeiten auch nach dem Rückstoßprinzip. Sie führen den Sauerstoff mit bzw. verwenden Treibstoffe, die auch den notwendigen Sauerstoff enthalten. Raketenantriebe arbeiten daher auch im luftleeren Raum.

2.6.5.8. Joule-Thomson-Effekt. Die innere Energie eines realen Gases ist volumenabhängig. Daher tritt bei der gedrosselten Entspannung (z.B. Gay-Lussac-Versuch) im Unterschied zum idealen Gas eine Temperaturänderung auf. Abb. 2.6.5–11 zeigt die gedrosselte Entspannung eines wärmeisolierten Systems vom Druck p_1 auf den Druck p_2. Die Änderung der inneren Energie $U = U_1 + U_2$ des Gesamtsystemes wird durch die über den linken Kolben zugeführte und über den rechten Kolben abgeführte Arbeit bestimmt, $dU_1 + dU_2 = p_1 dV_1 - p_2 dV_2$. Man beachte $dV_1 < 0$, $dV_2 > 0$. Man kann auch sagen, die Gesamtenthalpie $H = H_1 + H_2$ bleibt konstant, $H_1 = U_1 + p_1 V_1$ bzw. $H_2 = U_2 + p_2 V_2$. Das Gas hat also vor dem Entspannen die gleiche Enthalpie wie nach dem Entspannen $H(p_1, T_1) = H(p_2, T_2)$. Für infinitesimale Druck- und Temperaturunterschiede gilt $dT = -(1/C_p)\,(\partial H/\partial p)_T dp \equiv \mu\, dp$ mit $dp < 0$ beim Entspannen. μ wird als *Joule-Thomson-Koeffizient* bezeichnet. Das Vorzeichen der Temperaturänderung wird durch das von $(\partial H/\partial p)_T$ bestimmt. Für das ideale Gas ist $H = C_p T$, man erhält $dT = 0$. Für das van-der-Waals-Gas gilt genähert $H = C_p T - [2a^2/(RT) - b]p$ und damit $\mu = [2a^2/(RT) - b]/C_p$. Es können beide Vorzeichen auftreten. Der Vorzeichenwechsel tritt bei der *Inversionstemperatur* $T_{JT} = 2a^2/Rb = (27/4)T_{kr}$ ein. Am Beispiel des van-der-Waals-Gases sieht man, daß die Abkühlung ($\mu > 0$) dadurch eintritt, daß gegen die zwischenatomaren Kräfte Arbeit geleistet werden muß, die zu Lasten der kinetischen Energie der Atome geht.

Abb. 2.6.5–12 zeigt den Joule-Thomson-Koeffizienten von Luft. Die Inversionstemperatur hat – bezogen auf die kritischen Daten – für alle Gase etwa denselben Verlauf, so wie für Stickstoff in Abb. 2.6.5–13. Wasserstoff und Helium fallen etwas heraus.

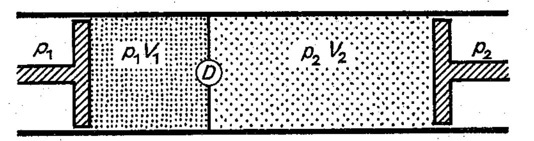


Abb. 2.6.5–11 Gedrosselte Entspannung

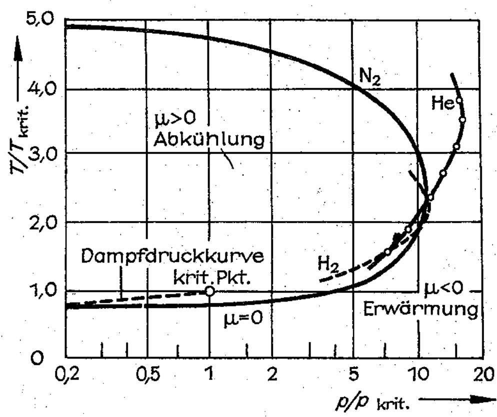


Abb. 2.6.5–13 Inversionstemperatur

2.6.5.9. Kälteerzeugung. Hier sollen nur einige technisch oder auch im Haushalt genutzte Möglichkeiten, ein System abzukühlen, besprochen werden. Erzeugung tiefer Temperaturen vgl. Kapitel 7.

Unter der Kälteerzeugung soll nicht die Abkühlung eines Systems mit einem vorhandenen Kältemittel verstanden werden, sondern die Herstellung des Kältemittels über thermodynamische Prozesse. Zum Verdampfen wird Verdampfungswärme, zum Schmelzen Schmelzwärme benötigt. Diese Energie wird der flüssigen (oder festen) Phase entzogen, die sich dadurch abkühlt. Gleichzeitig kühlt sich dabei die Umgebung ab, mit der die Flüssigkeit im Wärmekontakt steht. Dies ist die einfachste Weise, eine Abkühlung zu erreichen.

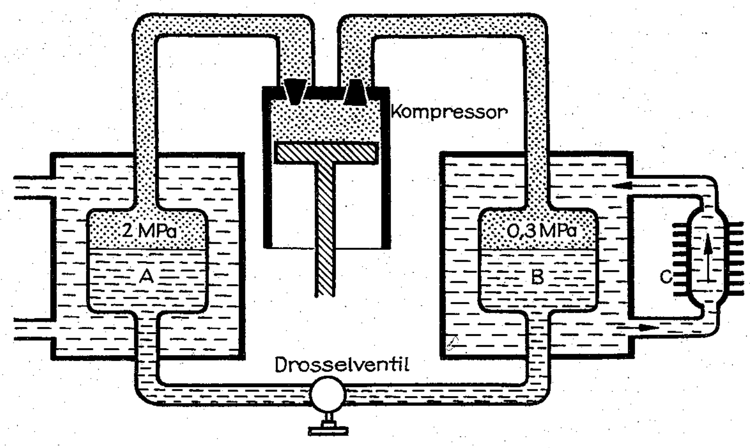


Abb. 2.6.5–14 Schematische Darstellung einer Kompressorkühlanlage

Das Lösen von Salzen in einer Flüssigkeit kann exotherm oder endotherm erfolgen. Im zweiten Fall wird die Lösungswärme der Flüssigkeit entzogen, sie kühlt sich ab.
In der *Kältemischung*, z.B. Eis und Kochsalz, werden beide Prozesse ausgenutzt. Wegen der Schmelzpunkterniedrigung in einer Lösung (vgl. 2.6.8.3.) schmilzt das Eis, das Salz löst sich im Wasser. Beide Prozesse erfordern Energie, Schmelzwärme und Lösungswärme. Sie wird der Kältemischung selbst entzogen, die sich dadurch abkühlt. Im genannten Fall können Temperaturen bis zu −20 °C erreicht werden.
Die *Abkühlung durch Verdampfen* wird in Kühlanlagen, z.B. im Haushaltkühlschrank, genutzt. Abb. 2.6.5–14 veranschaulicht eine mit Ammoniak als Arbeitsmittel betriebene Kühlanlage. Der Dampf wird aus B laufend abgesaugt. Mittels eines Kompressors und durch Kühlung (Außenluft oder Kühlwasser) wird der Dampf in A wieder verflüssigt und über ein Drosselventil zurückgeführt. Das Arbeitsmittel ist von einem Kühlmittel umgeben, das sich durch den Wärmekontakt mit dem Arbeitsmittel abkühlt. Das Kühlmittel wird dann über ein Rohrsystem zu den zu kühlenden Teilen C geführt. In Haushaltkühlschränken wird direkt mit dem Arbeitsmittel (Freon) gekühlt.
In der *Lindeschen Kältemaschine* wird der Joule-Thomson-Effekt z.B. zur Luftverflüssigung genutzt. Die Volumenabhängigkeit der inneren Energie realer Gase führt dazu, daß sich das Gas unterhalb der Inversionstemperatur bei einer gedrosselten Entspannung abkühlt. Wie Abb. 2.6.5–15 zeigt, wird Außenluft angesaugt und durch einen Kompressor auf z. B. 20 MPa verdichtet. Der Luft werden Wasser und Kohlendioxid entzogen, die Kompressionswärme wird durch das Kühlwasser abgeführt. Am Ende eines Rohres befindet sich ein Drosselventil, wo die Luft entspannt wird,

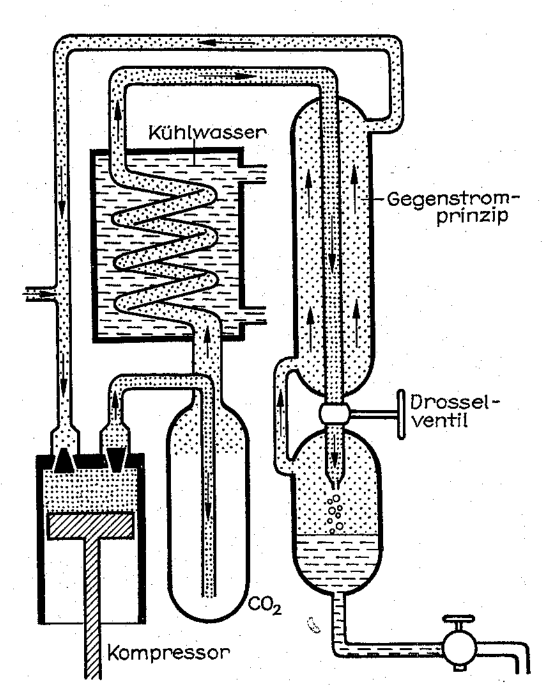


Abb. 2.6.5–15 Luftverflüssigung nach LINDE

z. B. auf Normaldruck 0,1 MPa. Bei dieser Entspannung kühlt sie sich aufgrund des Joule-Thomson-Effektes ab. Die kalte Luft kühlt im Gegenstromprinzip die im Rohr zugeführte Luft vor, bis deren Temperatur so niedrig liegt, daß die Abkühlung bei der Entspannung zur Verflüssigung führt. Die flüssige Luft kühlt sich durch Verdampfen weiter ab, bis die zum Druck des entspannten Gases gehörende Sättigungstemperatur (−190 °C bei Normaldruck) erreicht ist. Beim Verdampfen der Luft entweicht mehr Stickstoff als Sauerstoff, da der Siedepunkt von Stickstoff tiefer liegt. Daher enthält die flüssige Luft etwa 50 % Sauerstoff.
Beim Entspannen des Gases tritt eine Abkühlung nur auf, wenn dessen Temperatur unterhalb der Inversionstemperatur liegt, bei höheren Temperaturen erwärmt es sich. Soll z. B. Wasserstoff verflüssigt werden, dann muß er erst mit flüssiger Luft vorgekühlt werden.

2.6.6. Phasengleichgewichte

Prozesse wie Schmelzen, Verdampfen, Züchten von Kristallen, chemische Reaktionen u. a. sind dadurch gekennzeichnet, daß mehrere *Phasen* nebeneinander bestehen und daß in dem thermodynamischen System mehrere Komponenten enthalten sind. Wegen der Vielfalt der möglichen Erscheinungen ist es besonders wichtig, die allgemeinen Aussagen über das thermodynamische Gleichgewicht solcher Systeme zu kennen.

2.6.6.1. Molare Größen. Für die Kennzeichnung des thermodynamischen Gleichgewichtes ist die Größe des Gesamtsystems unwesentlich. Wenn alle extensiven Größen verdoppelt oder halbiert werden, laufen die Prozesse in gleicher Weise ab. Daher werden zur Vereinfachung der Beschreibung molare, d. h. auf das Mol bezogene, extensive Größen eingeführt. Schreibt man die Molzahl z in den thermodynamischen Potentialen explizit als Variable, dann gilt nach dem Gesagten $G(p, T, z) = z\,G(p, T, 1)$, $F(V, T, z) = z\,F(V/z, T, 1)$ usw. Als *molare freie Enthalpie* bzw. molare freie Energie werden $g(p, T) = G(p, T, 1)$ bzw. $f(V/z, T) = F(V/z, T, 1)$ bezeichnet. Die freie Enthalpie nimmt unter den thermodynamischen Potentialen dadurch eine Sonderstellung ein, daß sie von den intensiven Zustandsgrößen p, T abhängt, während in allen anderen Fällen zumindest eine extensive Zustandsgröße als Variable auftritt. Daher gilt $g(p, T) = [\partial G(p, T, z)/\partial z]_{p,T}$, während z. B. $f \neq [\partial F(V, T, z)/\partial z]_{V,T}$ ist.
In einem *einkomponentigen System*, das aus *zwei Phasen* I und II besteht (Abb. 2.6.6–1a), ist die freie Enthalpie

$$G(p, T, z^{\mathrm{I}}, z^{\mathrm{II}}) = z^{\mathrm{I}} g^{\mathrm{I}}(p, T) + z^{\mathrm{II}} g^{\mathrm{II}}(p, T).$$

Die molaren freien Enthalpien der beiden Phasen sind unterschiedliche Funktionen von Druck und Temperatur. Die gesamte freie Enthalpie ist die Summe der freien Enthalpien der beiden Phasen.
Enthält ein *einphasiges System zwei Komponenten* (Abb. 2.6.6–1b), dann führt man über $g_1(p, T, x) = \partial G/\partial z_1$ und $g_2(p, T, x) = \partial G/\partial z_2$ die *partiellen, molaren freien Enthalpien* g_i der Komponenten i ein. Sie hängen noch von der Konzentration x ab. $x = z_2/(z_1 + z_2)$ kann zwischen Null und Eins variieren. Wegen der intensiven Variablen in der freien Enthalpie gilt wiederum

$$G(p, T, z_1, z_2) = z_1 g_1(p, T, x) + z_2 g_2(p, T, x).$$

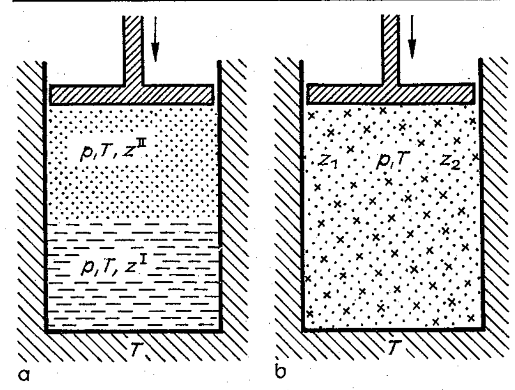


Abb. 2.6.6–1 (a) zwei Phasen eines einkomponentigen Systems, (b) zwei Komponenten in einer Phase

Für ein *zweikomponentiges, zweiphasiges System* ergibt sich $G(p, T, z_1^{\mathrm{I}}, z_1^{\mathrm{II}}, z_2^{\mathrm{I}}, z_2^{\mathrm{II}}) = z_1^{\mathrm{I}} g_1^{\mathrm{I}}(p, T, x^{\mathrm{I}}) + z_2^{\mathrm{I}} g_2^{\mathrm{I}}(p, T, x^{\mathrm{I}}) + z_1^{\mathrm{II}} g_1^{\mathrm{II}}(p, T, x^{\mathrm{II}}) + z_2^{\mathrm{II}} g_2^{\mathrm{II}}(p, T, x^{\mathrm{II}})$. Entsprechend kompliziertere Ausdrücke erhält man bei mehreren Komponenten und mehreren Phasen. Von der Komponente 1 sind insgesamt $z_1 = z_1^{\mathrm{I}} + z_1^{\mathrm{II}}$ Mol vorhanden, analog dazu ist $z_2 = z_2^{\mathrm{I}} + z_2^{\mathrm{II}}$.

2.6.6.2. Chemisches Potential. Betrachtet man eine Zustandsänderung in einem homogenen, d. h. einphasigen System mit K Komponenten, so verändern sich die thermodynamischen Potentiale, z. B. die freie Enthalpie $G = U - TS + pV$, bei einer Veränderung der Molzahlen entsprechend

$$dG = S dT + V dp + \sum_k \mu_k dz_k.$$

Analog gelten für die innere bzw. freie Energie $dU = T dS - p dV + \sum_k \mu_k dz_k$, $dF = S dT - p dV + \sum_k \mu_k dz_k$. Die auftretenden Funktionen μ_k werden als chemische Potentiale (der Komponente k) bezeichnet. Ihr Name weist darauf hin, daß sie den Ablauf chemischer Reaktionen (bei denen sich ja die Molzahlen ändern) maßgeblich bestimmen. Der Definition entsprechend sind sie aus $\mu_k = (\partial G/\partial z)_{p,T}$ berechenbar. Daraus folgt $\mu_k = g_k$. Innerhalb einer Phase ist für jede Komponente das chemische Potential gleich der molaren freien Enthalpie.

2.6.6.3. Temperatur und Druckausgleich. Ein abgeschlossenes System kann aus Teilsystemen mit unterschiedlicher Temperatur und unterschiedlichem Druck bestehen. Werden ein Wärmetransport durch die Trennwände und ein Verschieben der Trennwände zugelassen, dann gleichen sich erfahrungsgemäß Temperatur und Druck aus. Aus den Hauptsätzen der Wärmelehre kann man auch diese Aussagen ableiten, denn die Hauptsätze bestimmen die Form des thermodynamischen Gleichgewichtszustandes. In einem abgeschlossenen System ist der Gleichgewichtszustand erreicht, wenn die Entropie ihren Maximalwert angenommen hat. Dabei können nur solche Prozesse ablaufen, bei denen Gesamtvolumen $V = \sum_i V^{(i)}$ und gesamte innere Energie $U = \sum_i U^{(i)}$ unverändert bleiben. Die Variationsforderung $\delta(S - \alpha U - \beta V) = 0$ führt dann mit $dS^{(i)} = (dU^{(i)} + p^{(i)} dV^{(i)})/T^{(i)}$ auf $T^{(i)} = T$ und $p^{(i)} = p$, wenn man die Variationsparameter in der Form $\alpha = 1/T$ bzw. $\beta = p/T$ schreibt.

Die Entropie hat einen Extremwert erreicht, nachdem Temperatur und Druckausgleich stattgefunden haben. Damit dieser Extremwert ein Maximum ist, muß u. a. $\partial p/\partial V < 0$ gelten, d. h., die Kompressibilität muß positiv sein.

2.6.6.4. Allgemeine Gleichgewichtsbedingung. Oft liegt der Fall vor, daß ein thermodynamisches System nicht abgeschlossen ist, daß aber Druck und Temperatur vorgegebene Werte aufweisen. Es interessiert dann, welche Phasen im thermodynamischen Gleichgewicht auftreten und wie groß die Konzentration der einzelnen Komponenten in diesen Phasen ist. Ein Beispiel ist das Erschmelzen einer Legierung in einem Ofen unter Atmosphärendruck. Der thermodynamische Gleichgewichtszustand ist dadurch bestimmt, daß die freie Enthalpie einen Minimalwert angenommen hat (vgl. 2.6.4.3.). Nachdem Druck und Temperatur in allen Phasen gleich sind, kann sich die freie Enthalpie nur dadurch ändern, daß sich die Konzentrationen der einzelnen Komponenten verschieben (Abb. 2.6.6–2), indem zwischen den Phasen ein Austausch von Atomen oder Molekülen stattfindet. Die freie Enthalpie ändert sich bei einem solchen Teilchenaustausch, solange das chemische Potential dieser Komponente in den beiden Phasen noch einen unterschiedlichen Wert hat. Der thermodynamische Gleichgewichtszustand ist dadurch gekennzeichnet, daß das chemische Potential (oder die molare freie Enthalpie) einer Komponente in allen Phasen (i, j) den gleichen Wert aufweist, $g_k^{(i)} = g_k^{(j)}$. Diese Aussage gilt in gleicher Weise für alle Komponenten k. Es ist jetzt möglich, die *Anzahl der Freiheitsgrade* eines thermodynamischen Systems, das aus K Komponenten und L Phasen besteht, zu bestimmen. Wird einem System eine neue Komponente hinzugefügt, dann erhöht sich die Anzahl der Freiheitsgrade dadurch, daß in jeder Phase eine neue Konzentration als

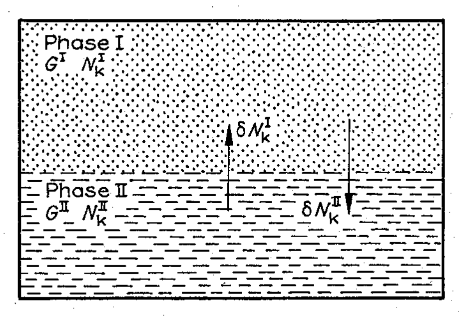


Abb. 2.6.6–2 Teilchenaustausch

Parameter auftritt. Wird die Anzahl der Phasen erhöht, dann verringert sich die Anzahl der Freiheitsgrade dadurch, daß durch die Gleichgewichtsbedingung $g_k^{(i)} = g_k^{(j)}$ zusätzliche Nebenbedingungen zu erfüllen sind. Die Anzahl der Freiheitsgrade erhöht sich, da in der zusätzlichen Phase neue Konzentrationen als Parameter auftreten.

Die Bilanz der Freiheitsgrade ist in der *Gibbsschen Phasenregel* zusammengefaßt. Die Anzahl der Freiheitsgrade beträgt

$$f = K - L + 2.$$

Ein einphasiges, einkomponentiges System hat also zwei Freiheitsgrade (p, T). In dieser Bilanz ist die denkbare Variation äußerer (elektrischer oder magnetischer) Felder nicht mit erfaßt. Sie kann die Anzahl der Freiheitsgrade noch erhöhen.

2.6.7. Phasengleichgewicht im einkomponentigen System

Unter bestimmten Bedingungen können mehrere Phasen nebeneinander bestehen, so z. B. unter Normaldruck Wasser und Wasserdampf bei 100 °C oder Wasser und Eis bei 0 °C. Diese Bedingungen können aus dem Phasendiagramm abgelesen werden. Formelmäßig sind sie, falls Druck und Temperatur festgelegt sind, durch die Gleichheit der molaren freien Enthalpien gegeben.

2.6.7.1. p-T-Diagramm. Abb. 2.6.7-1 zeigt das Phasendiagramm von CO_2 mit dem Existenzgebiet von fester, flüssiger und dampfförmiger Phase im thermodynamischen Gleichgewicht. Auf der *Dampfdruckkurve*, der *Schmelzdruckkurve* und der *Sublimationsdruckkurve* können jeweils zwei Phasen nebeneinander bestehen. Entsprechend der Gibbsschen Phasenregel bilden diese Zustände eine Mannigfaltigkeit mit einem Freiheitsgrad. Sollen alle drei Phasen nebeneinander existieren, dann verbleibt kein Freiheitsgrad, dies ist nur im sogenannten Tripelpunkt – für CO_2 bei −56,57 °C und 0,51 MPa – möglich. Unterhalb der zugehörigen Temperatur bzw. des zugehörigen Druckes kann die Dampfphase direkt in die feste Phase überführt werden, während

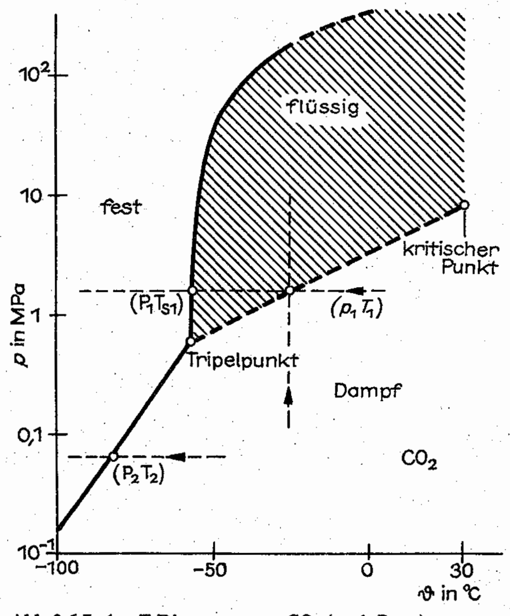


Abb. 2.6.7-1 p-T-Diagramm von CO_2 (nach POHL)

Tab. 2.6.7-2 Kritische Punkte

Gas	T_{kr} in °C	p_{kr}/p_0*	ϱ_{kr} in g/cm³
He	−267,9	2,26	0,069
H_2	−239,9	12,8	0,030
O_2	−118,4	50,1	0,38
H_2O	374,1	218,3	0,315
CO_2	31,1	73,0	0,465
NH_3	132,3	111,3	0,235

* $p_0 = 1{,}013\,25 \cdot 10^5$ Pa = Normaldruck

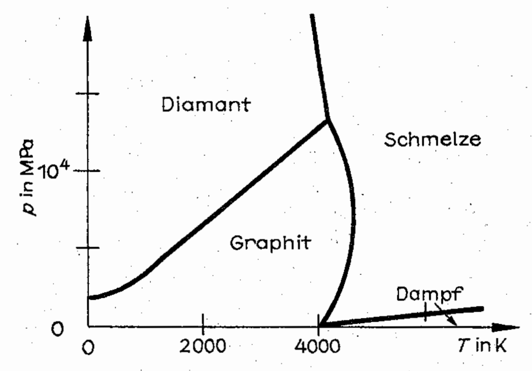


Abb. 2.6.7-3 p-T-Diagramm von Kohlenstoff (nach LANDOLT-BÖRNSTEIN)

für höher liegende p- bzw. T-Werte zunächst Verflüssigung eintritt. Hervorzuheben ist weiterhin die Existenz eines *kritischen Punktes* (Tab. 2.6.7-2). Oberhalb dieses Punktes kann zwischen Dampf und Flüssigkeit nicht mehr unterschieden werden. Zieht man zur qualitativen Diskussion die van-der-Waals-Gleichung $(p + z^2 a^2/V^2)(V - zb) = zRT$ heran, dann berechnet sich der kritische Punkt zu $V_{kr} = 3zb$, $T_{kr} = 8a^2/27bR$ und $p_{kr} = a^2/27b^2$. Danach gilt also $p_{kr}V_{kr}/zRT_{kr} = 8/3$. Die Meßwerte erfüllen diese Bedingungen aber nur mit einem Fehler von 20 bis 30%, was auf den Modellcharakter der van-der-Waals-Gleichung hinweist.

Wird Dampf beim Druck p_1 abgekühlt, dann erreicht er bei T_1 die Dampfdruckkurve. p_1 ist der zur Temperatur T_1 gehörige Dampfdruck. In diesem Zustand können Dampf und Flüssigkeit miteinander im Gleichgewicht stehen. Dampf in diesem Zustand nennt man *gesättigten Dampf.* Beim Versuch, die Temperatur weiter zu erniedrigen, verflüssigt sich zunächst der gesamte Dampf, dann tritt eine weitere Abkühlung ein. Bei der Temperatur T_{S1} ist die Schmelzdruckkurve erreicht. Flüssigkeit und feste Phase stehen miteinander im Gleichgewicht. Bei weiterer Abkühlung ist die gesamte Flüssigkeit erstarrt.

Wird Dampf abgekühlt, der unter dem Druck p_2 steht, dann wird bei der Temperatur T_2 die Sublimationsdruckkurve erreicht. Beim Versuch, weiter abzukühlen, geht der gesättigte Dampf direkt in die feste Phase über. Bei einer isothermen Kompression bei der Temperatur T_1 tritt beim Druck p_1 Verflüssigung ein. Es bildet sich eine Phasengrenze zwischen dampfförmiger und flüssiger Phase, bis der gesamte Dampf kondensiert ist. Bei einer isothermen Kompression oberhalb der kritischen Temperatur T_{kr} tritt keine Phasengrenze auf. Eine klare Unterscheidung zwischen Dampf und Flüssigkeit ist nicht möglich. Beide Phasen werden ja nur durch verschiedene Dichtewerte unterschieden, nicht durch ihre Struktur. Beides sind homogene Phasen mit sich regellos bewegenden Atomen oder Molekülen. Auf der Schmelzdruckkurve kann es keinen kritischen Punkt geben. Die feste Phase unterscheidet sich durch ihren kristallinen Aufbau mit der regelmäßigen Anordnung von Atomen an den Gitterplätzen prinzipiell in ihrer Struktur von der flüssigen und gasförmigen Phase. Eine Substanz kann verschiedene feste Phasen aufweisen (Abb. 2.6.7-3), die sich in ihrem kristallinen Aufbau unterscheiden.

Die Überführung in einen anderen Aggregatzustand ist mit einer Wärmetönung verbunden, ohne daß eine Temperatur-

Tab. 2.6.7–4 Umwandlungswärmen beim Schmelzen und Verdampfen

Substanz	T_s in°C	ΔH_s in $kJmol^{-1}$	T_v in °C	ΔH_v in $kJmol^{-1}$
Ag	961,93	11,3	22,10	252
Al	660,1	11,3	2480	251
Au	1064,43	12,8	2600	344
C[1]	3550	10,9	4200	715
Cu	1083	13,0	2600	307
Fe	1534	14,4	3000	341
Hg	−38,87	2,3	357	59,2
Mg	649	8,8	1103	128
Ni	1453	16,9	2900	374
Pb	327,4	4,81	1740	178
Pt	1769	21,8	4300	469
H_2O	0	6,0	100	40,6
C_2H_6OH	−114,4	5,0	78,3	40,3
CCl_4	−22,9	2,6	76,6	29,6
He	−272		−268	1,8
H_2	−259		−252	0,9
N_2	−210		−195	5,6
O_2	−218		−183	6,8
CO_2			−78,5[2]	25,2[2]

[1] Graphit, [2] Sublimation

änderung auftritt. Um 1 mol Eis von 0 °C in Wasser von 0 °C zu überführen, wird die *Schmelzwärme* 6,0 kJ/mol benötigt. Beim umgekehrten Prozeß muß eine *Erstarrungswärme* vom gleichen Betrag entzogen werden. Um eine Flüssigkeit am Siedepunkt in Dampf zu überführen, muß die *Verdampfungswärme* zugeführt werden, beim Verflüssigen wird *Kondensationswärme* abgegeben. Bei einem direkten Übergang vom festen in den dampfförmigen Aggregatzustand spricht man von *Sublimationswärme*, allgemein von *Umwandlungswärme* (Tab. 2.6.7–4). Erfolgen Schmelzen bzw. Verdampfen unter konstantem Druck, dann ist die Umwandlungswärme durch die Differenz der Enthalpien der beiden Phasen gegeben.

Bei der mathematischen Beschreibung nimmt man die Umwandlungswärme als einen deltafunktionsartigen Beitrag in die Temperaturabhängigkeit der spezifischen Wärme mit auf, $C_p(T) = C_{p1}(T)\,\theta(T_0 - T) + \Delta H\delta(T - T_0) + C_{p2}(T)\,\theta(T - T_0)$. Die Umwandlungstemperatur T_0 für den Übergang von der Phase 1 in die Phase 2 kann nur überschritten werden, wenn dem System die Umwandlungswärme zugeführt wird. Da die Molwärme verschiedener Phasen eine unterschiedliche Temperaturabhängigkeit hat, wurden zwei Funktionen C_{p1} und C_{p2} eingeführt und durch die Stufenfunktion $\theta(x) = 0$ für $x < 0$, $\theta(x) = 1$ für $x > 0$ auf den jeweiligen Temperaturbereich beschränkt.

2.6.7.2. Dampfdruckkurve. Ein einkomponentiges, einphasiges System hat zwei Freiheitsgrade, z. B. p und T. Zwei Phasen können nur auf der Dampfdruckkurve (oder Schmelzdruckkurve usw.) miteinander im Gleichgewicht stehen. Das einkomponentige, zweiphasige System hat nur noch einen Freiheitsgrad. Die einschränkende Bedingung besteht darin (vgl. 2.6.6.4.), daß die molaren freien Enthalpien g^i gleich groß sein müssen, $g^{\text{flüss.}} = g^{\text{gasf.}}$. Diese Bedingung liefert die Dampfdruckkurve. Um die Änderung des Sättigungsdruckes mit der Temperatur auf andere Meßgrößen zurückführen zu können, betrachtet man eine infinitesimale Zustandsänderung auf der Dampfdruckkurve (Abb. 2.6.7–5a). Die molare freie Enthalpie ändert sich dabei entsprechend $\mathrm{d}g = S\mathrm{d}T + V\mathrm{d}p$. Aus $\mathrm{d}g^{\text{gasf.}} = \mathrm{d}g^{\text{flüss.}}$ folgt die *Clausius-Clapeyronsche Dampfdruckformel*

$$\frac{\mathrm{d}p}{\mathrm{d}T} = \frac{Q_V}{T(V^{\text{gasf.}} - V^{\text{flüss.}})},$$

wenn man die Entropieänderung beim (isothermen) Verdampfen $S^{\text{gasf.}} - S^{\text{flüss.}} = Q_V/T$ durch die Verdampfungswärme ausdrückt. Tab. 2.6.7–6 gibt an, wie sich die Siedetemperatur von Wasser mit dem Druck verschiebt. Die Clausius-Clapeyronsche Formel gilt für alle Phasengleich-

Tab. 2.6.7–6 Dampfdruck des gesättigten Wasserdampfes bzw. Siedetemperatur für verschiedene Drücke

ϑ in °C	p/p_0	ϑ in °C	p/p_0
0	0,00602	60	0,1965
10	0,01211	70	0,3075
20	0,02307	80	0,4672
30	0,04187	90	0,6917
40	0,07279	100	1,0000
50	0,12172	110	1,413

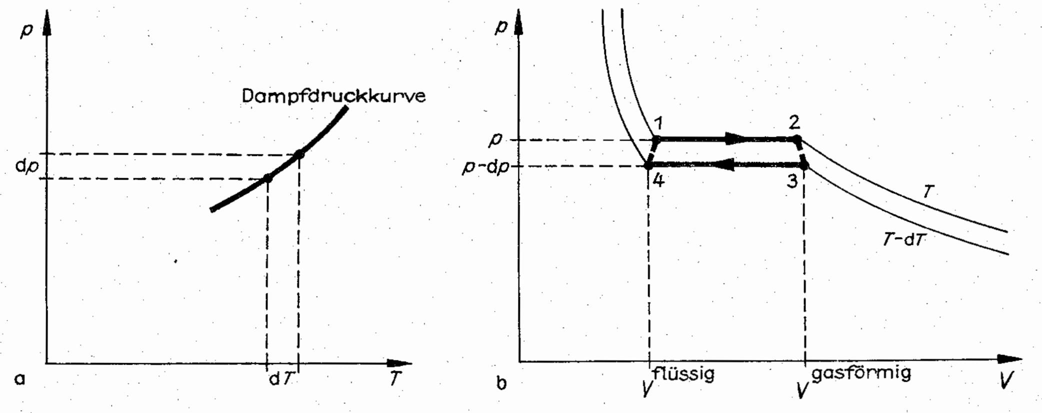


Abb. 2.6.7–5 (a) Zustandsänderung entlang der Dampfdruckkurve, (b) Kreisprozeß zur Bestimmung der Dampfdruckformel

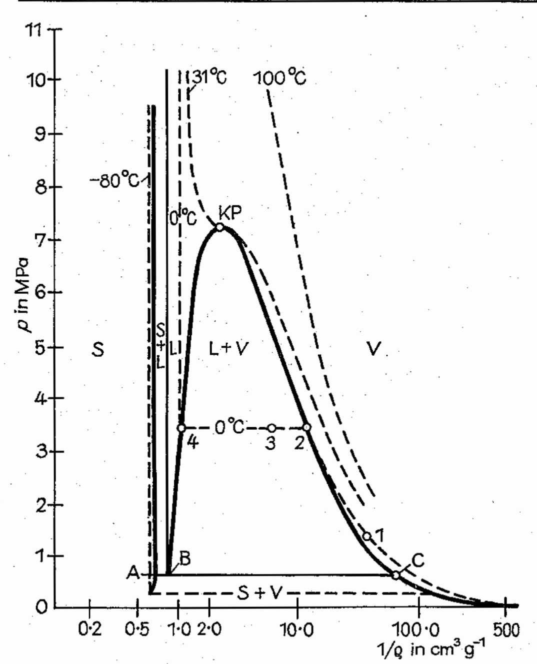


Abb. 2.6.7–7 p-V-Diagramm des CO_2, S feste, L flüssige, V dampfförmige Phase. Am Tripelpunkt KP haben die drei Phasen die durch die Punkte A, B und C gekennzeichneten Dichten (nach KING)

gewichte, nur daß Q_V und ΔV durch die zu den beiden betrachteten Phasen gehörenden Werte zu ersetzen sind.

2.6.7.3. *p-V*-Diagramm. Abb. 2.6.7–7 zeigt das p-V-Diagramm von CO_2. Zur Diskussion der auftretenden Phasen werden Isothermen T = konst. eingetragen. Sie ähneln nur oberhalb der kritischen Temperatur den Isothermen des idealen Gases, werden aber durch die van-der-Waals-Gleichung (vgl. Abb. 2.6.1–5) qualitativ richtig beschrieben. Unterhalb der kritischen Temperatur T_{kr} gibt es ein Zweiphasengebiet. Von niedrigem Druck ausgehend (Punkt 1 der Isothermen zu 0 °C) steigt der Druck bei der isothermen Kompression zunächst an. Am Punkt 2 ist der zur betrachteten Temperatur gehörige Sättigungsdruck erreicht. Bei weiterer Volumenverminderung kondensiert ein Teil des Dampfes, ohne daß der Druck dabei ansteigt (Punkt 3). Am Punkt 4 ist der gesamte Dampf verflüssigt. Wegen der geringen Kompressibilität der Flüssigkeit tritt auch bei sehr hohen Drücken nur noch eine kleine Volumenverminderung ein.

Es ist instruktiv, daß man die *Clausius-Clapeyronsche Dampfdruckformel* auch gewinnen kann, indem man einen – im p-V-Diagramm übersichtlich darstellbaren – Kreisprozeß zwischen zwei benachbarten Isothermen betrachtet (vgl. Abb. 2.6.7–5b). Bei der isothermen Expansion von 1 nach 2 muß die Verdampfungswärme Q_V zugeführt werden, bei der isothermen Kompression von 3 nach 4 wird die Kondensationswärme entzogen. Da beide Prozesse auch isobar verlaufen, ist die abgegebene Arbeit, die vom Kreisprozeß umschlossene Fläche, $dW = dp\Delta V$. Der Wirkungsgrad $\eta = dW/Q_V$ dieses reversibel geführten Prozesses ist andererseits durch die Temperaturen T und $T - dT$ bestimmt (vgl. 2.6.3.3.), $\eta = (T - [T - dT])/T$. Der Vergleich beider Ausdrücke liefert $dp/dT = Q_V/(T\Delta V)$.

2.6.7.4. Phasenumwandlung verschiedener Ordnung. Unter bestimmten Bedingungen können verschiedene Phasen miteinander im thermodynamischen Gleichgewicht stehen, so z. B. flüssige und dampfförmige Phase auf der Dampfdruckkurve. Der Zusammenhang $p = p(T)$ wird durch die Bedingung bestimmt, daß die molaren freien Enthalpien $g(p, T)$ der beiden Phasen gleich groß sein müssen. Abb. 2.6.7–8 veranschaulicht dies (für eine Phasenumwandlung erster Ordnung), indem der Verlauf $g(p, T)$ zusammen mit der Dampfdruckkurve im p-T-Diagramm und der Isothermen im p-V-Diagramm qualitativ dargestellt wird. Am Punkte p_1, T_1 gilt $g^I = g^{II}$. In der Abb. 2.6.7–6 schneiden sich die Kurven g^I und g^{II}, sie haben einen Berührungspunkt erster Ordnung. Haben sie einen Berührungspunkt n-ter Ordnung, dann spricht man von einem Phasenübergang n-ter Ordnung.

Aus den Hauptsätzen der Thermodynamik folgt $S = -(\partial G/\partial T)_p$, $V = (\partial G/\partial P)_T$, $C_p = -T(\partial^2 G/\partial T^2)_p$, $\alpha = (\partial^2 G/\partial T \partial p)/V$, $\varkappa_T = -(\partial^2 G/\partial p^2)_T/V$. Daher ergeben sich am Umwandlungspunkt bestimmte Aussagen über Entropie S, Volumen V, Molwärme bei konstantem Druck C_p,

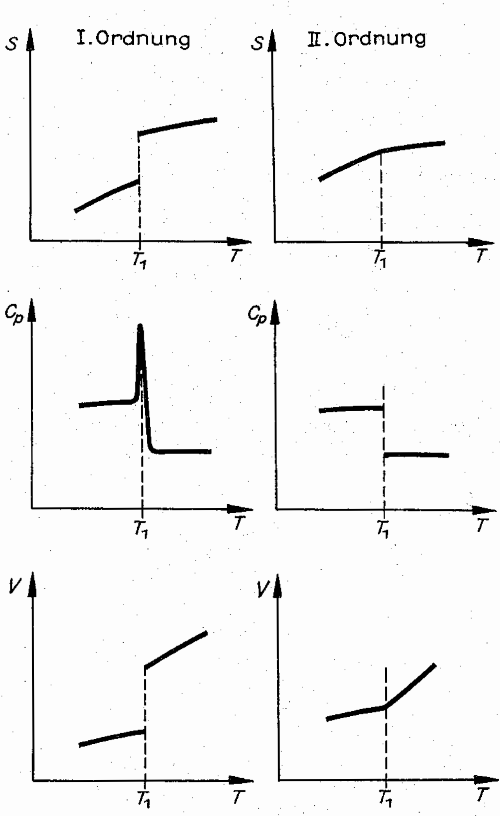


Abb. 2.6.7–9 Verlauf der Entropie, der Molwärme C_p und des Volumens in der Umgebung eines Umwandlungspunktes für eine Phasenumwandlung erster bzw. zweiter Ordnung

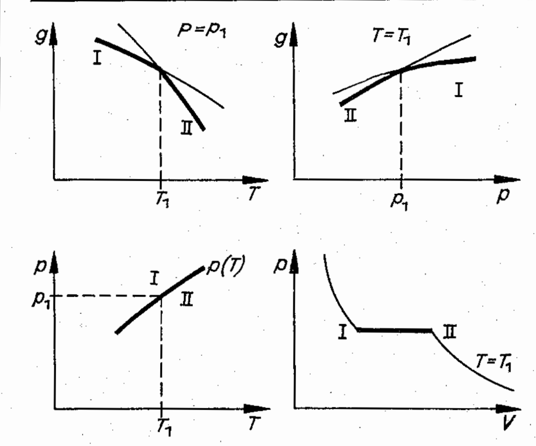


Abb. 2.6.7–8 Molare freie Enthalpie zweier Phasen I bzw. II in Zusammenhang mit Dampfdruckkurve und *p*-*V*-Diagramm

Ausdehnungskoeffizient α und isotherme Kompressibilität $\varkappa_T$. Abb. 2.6.7–9 zeigt qualitativ das Verhalten von S, C_p und V für einen Phasenübergang erster bzw. zweiter Ordnung.

Phasenübergang erster Ordnung. Hier tritt ein Sprung in der Entropie $\Delta S = S^{II} - S^{I}$ (und damit eine Umwandlungswärme) und im Volumen $\Delta V = V^{II} - V^{I}$ auf. Die Umwandlungskurve ist durch $dp/dT = \Delta S/\Delta V$ gegeben (Clausius-Clapeyronsche Formel). Die Molwärme enthält einen deltafunktionsartigen Anteil.

Phasenübergang zweiter Ordnung. Umwandlungswärme und Volumenänderung sind nicht vorhanden. Molwärme C_p wie auch α und $\varkappa_T$ ändern sich aber sprungartig. Die Umwandlungskurve ist durch $dp/dT = \Delta\alpha/\Delta\varkappa_T$ gegeben. Auf der Umwandlungskurve besteht der Zusammenhang $\Delta C_p \Delta\varkappa_T/(\Delta\alpha)^2 = VT$ zwischen den Sprüngen in den oben aufgeführten Größen.

Entsprechend können Phasenumwandlungen höherer Ordnung diskutiert werden. Es treten auch Übergangsfälle auf, z. B. wenn am Umwandlungspunkt die links- oder rechtsseitige Ableitung von g gegen Unendlich geht.

2.6.8. Phasengleichgewicht binärer Systeme

Ein binäres System hat drei Freiheitsgrade. Zu den zwei Freiheitsgraden (z. B. p, T), die vom einkomponentigen System bekannt sind, kommt die *Konzentration* hinzu. Sie wird in verschiedener Weise angegeben. Sind z_A und z_B die Molzahlen der beiden Komponenten, dann ist das *Molverhältnis* $x = z_B/(z_A + z_B)$. Es variiert zwischen $x = 0$ (reine Komponente A) und $x = 1$ (reine Komponente B). Der Wert $100x$ tritt bei der Angabe „Atomprozent Komponente B" auf. Bei der Angabe „Gewichtsprozent der Komponente B" wird das Verhältnis der Massenanteile $z_B M_B/(z_A M_A + z_B M_B)$ gebildet und mit 100 multipliziert. M_A und M_B sind die relativen Molekül- (oder Atom-)Massen. Viele andere Angaben sind zur Charakterisierung der Konzentration in Gebrauch, bei Lösungen z. B. die *Molarität* (Anzahl der gelösten Mol je Liter Lösung), bei in Flüssigkeiten absorbierten Gasen *Normkubikzentimeter je Gramm* (unter Normalbedingungen würde die je Gramm gelöste Gasmenge das entsprechende Volumen einnehmen).

2.6.8.1. Gemische von Gasen und Flüssigkeiten. Gase

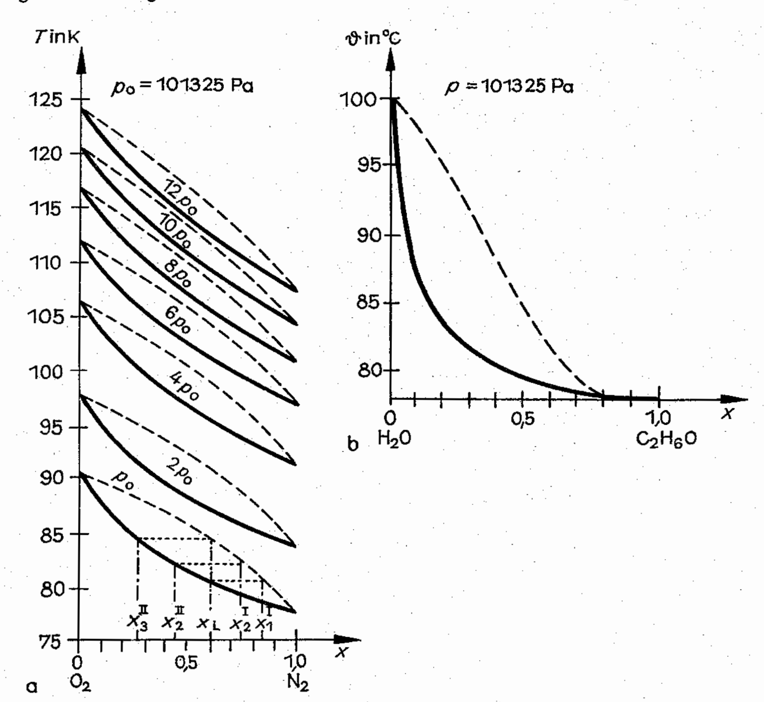


Abb. 2.6.8–1 Zustandsdiagramm binärer Systeme, (a) O_2-N_2 bei verschiedenen Drücken, (b) H_2O-C_2H_6O bei Normaldruck p_0 (nach LANDOLT-BÖRNSTEIN)

sind in beliebigen Konzentrationen mischbar. Für ideale Gase addieren sich einfach die *Partialdrücke* zum Gesamtdruck $p_A + p_B = p$. Die Zustandsgleichungen $p_A V = z_A RT$ und $p_B V = z_B RT$ führen auf die Zustandsgleichung $pV = zRT$ für das Gemisch.

Zwei Flüssigkeiten können unbeschränkt mischbar sein (Wasser und Ethanol), es kann jedoch auch eine Mischungslücke (vgl. 2.6.8.2.) auftreten (Wasser und Ether). Sind die Flüssigkeiten nicht mischbar (Wasser und Öl), dann kann höchstens eine Emulsion (Milch) hergestellt werden, indem – z. B. durch Ultraschalleinwirkung – das Öl in feinen Tröpfchen im Wasser verteilt wird.

Dem Zustandsdiagramm kann man entnehmen, welche Phasen im thermodynamischen Gleichgewicht bei gegebenen Werten von Druck, Temperatur und Konzentration vorhanden sind. Abb. 2.6.8–1a zeigt das Zustandsdiagramm des O_2-N_2-Systems, Abb. 2.6.8–1b das des Wasser-Ethanol-Systems. Da in der ebenen Darstellung nur zwei Parameter variiert werden können, das System jedoch drei Freiheitsgrade hat, wird z. B. der Druck vorgegeben – meist interessiert das Verhalten bei Normaldruck. Aus dem Zustandsdiagramm kann dann die Zusammensetzung des Systems in Abhängigkeit von Temperatur und Konzentration abgelesen werden.

Abb. 2.6.8–1 kann man entnehmen, daß bei Normaldruck Stickstoff bei 77,35 K und Sauerstoff bei 90,2 K siedet. Oberhalb der *Taulinie* ist nur die Dampfphase, unterhalb der *Siedelinie* nur die flüssige Phase vorhanden. Sauerstoff und Stickstoff sind unbeschränkt mischbar. Im Zwischenbereich bestehen Dampf und Flüssigkeit nebeneinander. Die Konzentration dieser beiden Phasen (x^{I} für den Dampf, x^{II} für die Flüssigkeit) kann voneinander verschieden sein. Sie wird durch die aus den Hauptsätzen der Thermodynamik folgenden Gleichgewichtsbedingungen bestimmt. x^{I} und x^{II} werden aus dem Zustandsdiagramm abgelesen, indem man die zur betrachteten Temperatur gehörenden

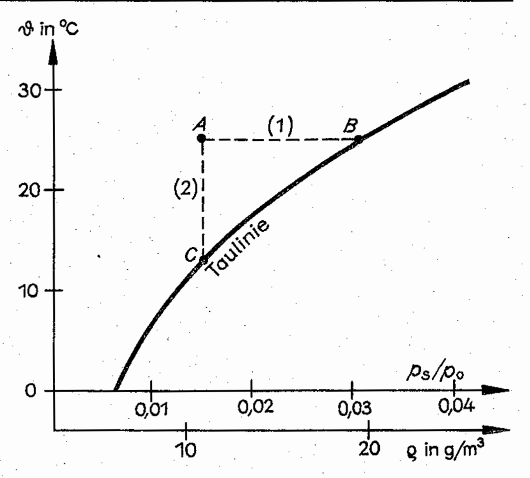


Abb. 2.6.8–2 Dampfspannung bzw. Dichte des gesättigten Wasserdampfes, $p_0 = 101\,325$ Pa

Punkte der Tau- bzw. Siedelinie aufsucht. Die Aufteilung der Gesamtmasse M in Anteile Flüssigkeit und Dampf berechnet sich nach dem *Hebelgesetz*

$M_{\mathrm{Fl}} = M(x - x^{\mathrm{I}})/(x^{\mathrm{II}} - x^{\mathrm{I}})$,

$M_{\mathrm{Dampf}} = M(x^{\mathrm{II}} - x)/(x^{\mathrm{II}} - x^{\mathrm{I}})$.

Wird ein System der Stickstoff-Konzentration x_L (vgl. Abb. 2.6.8–1) zunächst bei einer Temperatur unter der Siedelinie betrachtet, dann ist es flüssig (flüssige Luft). Beim Erwärmen wird die Siedelinie erreicht. Der Siedepunkt T_1 liegt über dem des reinen Stickstoffs und unter dem des reinen Sauerstoffs. Der entstehende Dampf hat die größere Stickstoffkonzentration x_1^{I}. Dadurch steigt die Sauerstoff-

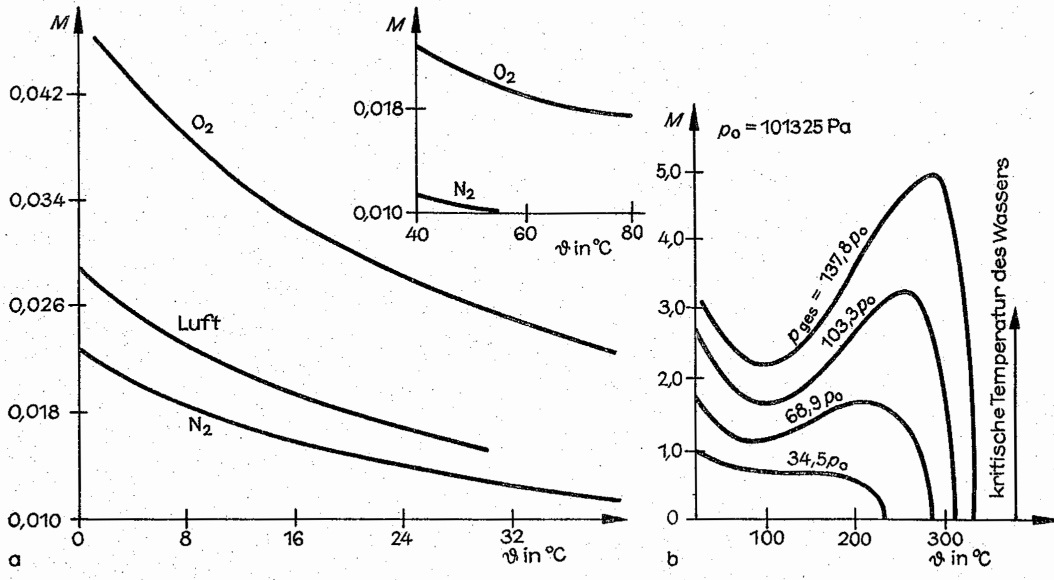


Abb. 2.6.8–3 (a) die in 1 g Wasser absorbierte Gasmenge M bei einem Partialdruck $p_0 = 101\,325$ Pa, (b) Druckabhängigkeit von M im Falle von O_2 (M in Normkubikzentimeter). $p_{\mathrm{ges.}}$ enthält auch den Dampfdruck des Wassers (nach LANDOLT-BÖRNSTEIN)

konzentration der Flüssigkeit über den Wert $x_1^{II} = x_L$ an. Bei einer etwas erhöhten Temperatur T_2 sind die Konzentrationen x_2^I und x_2^{II}. Wird der Taupunkt erreicht, dann verdampft der letzte Rest der Flüssigkeit, die hier die Konzentration x_3^{II} hat, während der Dampf wieder die Ausgangskonzentrationen $x_3^I = x_L$ erreicht hat. Man beachte, daß bei dem betrachteten Prozeß der Dampf nicht abgesaugt wird. Es liegt ein abgeschlossenes, veränderliches Volumen vor, in dem sich das System unter konstantem Druck befindet.

Luftfeuchtigkeit charakterisiert den Gehalt der Luft an Wasserdampf. Die *absolute Luftfeuchtigkeit* (Wassermasse/Volumen) ist gleich der Dichte des Wasserdampfes. Die feuchte Luft ist ein zweikomponentiges System, in dem die eine Komponente (Luft) in dem interessierenden Temperaturbereich als ideales Gas angesehen werden kann, während für die andere Komponente (Wasser) Dampf und Flüssigkeit nebeneinander bestehen können. Abb. 2.6.8–2 zeigt die Dampfspannung des gesättigten Wasserdampfes. Der Wassergehalt wird durch den Dampfdruck (Partialdruck des Wasserdampfes) oder die Dichte des Wassers gekennzeichnet. Erhöht man bei gegebener Temperatur den Wassergehalt der Luft (Kurve 1), dann wird beim Sättigungsdampfdruck (Punkt B) die *Taulinie* erreicht. Weiterer Wasserdampf, der der Luft zugeführt wird, kondensiert. Als *relative Luftfeuchtigkeit* bezeichnet man das Verhältnis der Dichte des Wasserdampfes der betrachteten Luft zur Dichte des Wasserdampfes gesättigter Luft der gleichen Temperatur. Sie beträgt auf der Taulinie 100%.

Wird Luft gegebenen Wassergehaltes abgekühlt (Kurve 2), dann erhöht sich die relative Luftfeuchtigkeit. Am Taupunkt (Punkt C in Abb. 2.6.8–2) beträgt sie 100%, da mit abnehmender Temperatur die Sättigung schon bei geringerer absoluter Luftfeuchtigkeit eintritt. Beim *Taupunkthygrometer* wird die Messung der Luftfeuchtigkeit auf eine Temperaturmessung zurückgeführt. Aus der Temperatur des Taupunktes kann auf die Feuchtigkeit bei der Ausgangstemperatur geschlossen werden. Das *Haarhygrometer* nutzt die Abhängigkeit der Spannung bzw. Länge bestimmter Fasern von Ihrem Wassergehalt, der wiederum durch die Feuchtigkeit der umgebenden Luft festgelegt wird. Im *Absorptionshygrometer* wird der Luft der Wasserdampf durch eine hygroskopische Substanz entzogen. Deren Massenzunahme dividiert durch das Luftvolumen ergibt die absolute Luftfeuchtigkeit.

Im Wasser ist auch eine gewisse Menge Luft absorbiert. Die in einer Flüssigkeit *absorbierte Gasmenge* hängt von Temperatur, Druck und Gasart ab. Sauerstoff löst sich in Wasser unter Normalbedingungen stärker als Stickstoff (Abb. 2.6.8–3a), so daß das Verhältnis der Mengen ≈ 2 : 1 beträgt, in Luft dagegen 1 : 4. Das ist für die Lebensbedingungen der Fische wesentlich. Mit der Erhöhung des Druckes steigt die gelöste Gasmenge (Abb. 2.6.8–3b), ihre Temperaturabhängigkeit kann nichtmonoton sein, unter Normaldruck nimmt beim Erwärmen der Luftgehalt des Wassers ab (Abb. 2.6.8–3a). Das Perlen von Selters oder Sekt ist dadurch bedingt, daß nach Öffnen der Flasche und der damit verbundenen Druckverminderung die Flüssigkeit übersättigt ist. Die gelöste Gasmenge ist in gewissen Grenzen – nicht in der Nähe des Taupunktes – proportional zum Partialdruck des Gases. Der Gesamtdruck ist nicht maßgebend. Man kann also z. B. das Entweichen von Kohlendioxid aus einer Selters nicht unterbinden, indem man den Luftdruck erhöht.

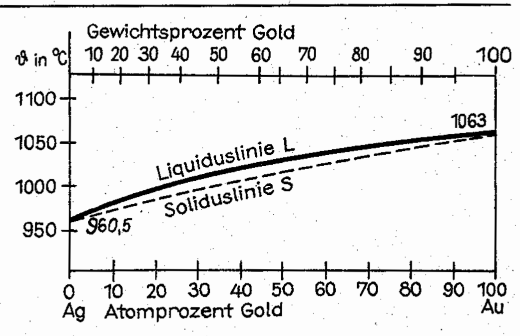


Abb. 2.6.8–4 Zustandsdiagramm bei lückenloser Mischbarkeit (nach HANSEN-ANDERKO)

2.6.8.2. Legierungen. Abb. 2.6.8–4 zeigt das Zustandsdiagramm des Ag-Au-Systems in dem Temperaturbereich, in dem Schmelze und Kristall miteinander im Gleichgewicht stehen. Silber und Gold sind *lückenlos mischbar.* Es tritt nur e i n e feste Phase auf, ein kubisch-flächenzentrierter (kfz) *Mischkristall,* in dem die Gitterplätze der Konzentration entsprechend mit Silber- oder Goldatomen besetzt sind. Wird beim Erwärmen eines Mischkristalles die *Soliduslinie* erreicht, dann beginnt die Ausbildung der flüssigen Phase. Diese hat eine Konzentration, die dem zu dieser Temperatur gehörenden Punkt auf der *Liquiduslinie* entspricht. Diese Aussagen sind analog zu denen, die aus Abb. 2.6.8–1 entnommen wurden.

Abb. 2.6.8–5 zeigt das Zustandsdiagramm des Ag-Cu-Systems. Es treten drei Phasen auf, zwei feste Phasen (α, β) und die Flüssigkeit (γ). Die Phase α hat die Struktur, in der das reine Silber kristallisiert (kfz). 960,5 °C ist der Schmelzpunkt des reinen Silbers. In dem *α-Mischkristall* befinden sich auf einem Teil der Gitterplätze Kupferatome. Das reine Kupfer schmilzt bei 1083 °C. Im *β-Mischkristall* ist ein Teil der Gitterplätze mit Silberatomen besetzt.

Die Gebiete $\alpha + \beta$, $\alpha + \gamma$ und $\beta + \gamma$ sind Zweiphasengebiete. Neben der Gesamtkonzentration x interessiert dann auch die Konzentration in den einzelnen Phasen x^α, x^β bzw.

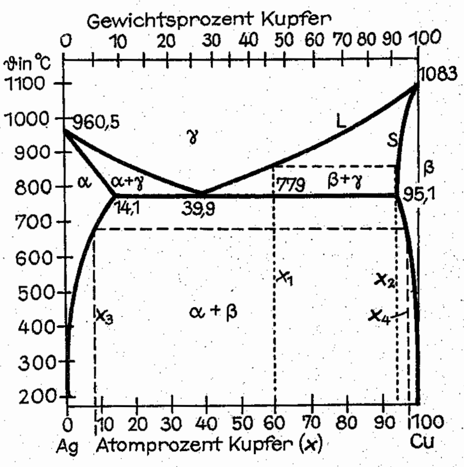


Abb. 2.6.8–5 Zustandsdiagramm mit Mischungslücke (nach HANSEN-ANDERKO)

x^γ. So stehen bei 850 °C in einem System mit der Konzentration $x_1 \leqq x \leqq x_2$ die Schmelze der Konzentration $x^\gamma = x_1$ und die feste Phase β der Konzentration $x^\beta = x_2$ miteinander im Gleichgewicht. Variiert man x zwischen x_1 und x_2, dann verschiebt sich nur das Mengenverhältnis der beiden Phasen entsprechend dem Hebelgesetz.

Bei 680 °C erhält man für Konzentrationen $x \leqq x_3$ einen α-Mischkristall mit der Konzentration $x^\alpha = x$, für Konzentrationen $x \geqq x_4$ einen β-Mischkristall mit der Konzentration $x^\beta = x$. Im Konzentrationsbereich $x_3 < x < x_4$ liegt eine Gemenge von α-Mischkristallen der Konzentration x_3 und β-Mischkristallen der Konzentration x_4 vor. Im Ag-Cu-System gibt es also eine *Mischungslücke.* Alle drei Phasen können nur in einem Punkt miteinander im thermodynamischen Gleichgewicht stehen, da entsprechend der Gibbsschen Phasenregel ein Freiheitsgrad weniger vorhanden ist. Im Ag-Cu-System tritt dies bei Normaldruck bei etwa 780 °C und 40 Atomprozent Kupfer auf.

Intermediäre Phasen sind in vielen Systemen zu finden, z. B. im Cu-Zn-System (Abb. 2.6.8–6). Die Struktur dieser Phasen ist meist von der der reinen Metalle verschieden. So ist der Kupfermischkristall (α-Phase) kubisch-flächenzentriert, die β-Phase kubisch-raumzentriert (krz), die γ-Phase hat eine kubische Struktur mit 52 Atomen je Elementarzelle, die η-Phase ist eine hexagonal dichteste Packung, der Zinkmischkristall (ε-Phase) hat ebenfalls eine hexagonal dichtest gepackte Struktur. Die intermediären Phasen haben einen relativ schmalen Existenzbereich und werden durch Mischungslücken voneinander getrennt.

Abb. 2.6.8–7 zeigt einen Ausschnitt aus dem Fe-C-Zustandsdiagramm mit dem α-Mischkristall (krz), dem γ-Mischkristall (kfz) zwischen 911 und 1392 °C, dem δ-Mischkristall (krz) mit dem Schmelzpunkt bei 1536 °C und dem (metastabilen) Zementit (Fe_3C). Beim reinen Eisen erfolgt am Curie-Punkt $T_c = 768$ °C eine Umwandlung vom ferromagnetischen zum paramagnetischen Eisen.

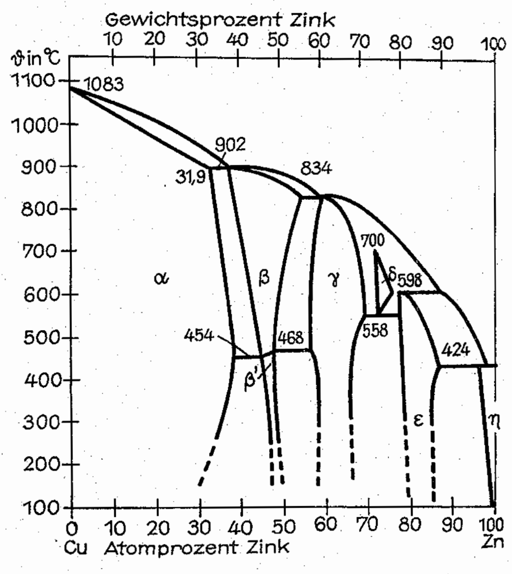


Abb. 2.6.8–6 Zustandsdiagramm mit intermediären Phasen (nach Hansen-Anderko)

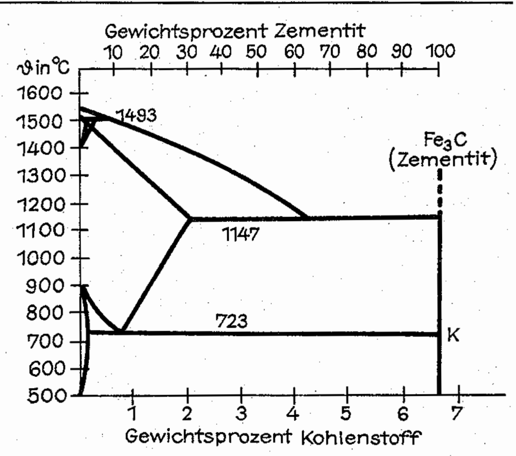


Abb. 2.6.8–7 Ausschnitt aus dem Fe-C-Zustandsdiagramm (nach Eisenkolb)

Das α-Eisen kann nur sehr wenig Kohlenstoff lösen, das γ-Eisen bis zu zwei Gewichtsprozent. Dies ist für die Herstellung von Stahl von Bedeutung. Wird kohlenstoffreiches γ-Eisen relativ rasch abgekühlt, dann kommt es nicht zum Ausscheiden von Zementit, weil die erforderliche Kohlenstoffdiffusion langsam vor sich geht. Es stellt sich ein metastabiler Zustand ein (martensitische Umwandlung), der sich durch eine hohe Festigkeit auszeichnet.

2.6.8.3. Lösungen. Im engeren Sinne versteht man unter einer Lösung eine flüssige Phase. Es wird ein Temperaturbereich zwischen Gefrierpunkt und Siedepunkt der einen Komponente, des *Lösungsmittels,* betrachtet. In diesem Temperaturbereich soll die reine zweite Komponente in fester Form vorliegen.

Die Löslichkeit gibt die Maximalkonzentration des Salzes im Lösungsmittel an, bei höherer Konzentration fallen Kristalle aus. Abb. 2.6.8–8 zeigt, daß die Löslichkeit temperaturabhängig ist. Bei verdünnten Lösungen beobachtet man eine der Konzentration proportionale Siedepunktserhöhung bzw. Gefrierpunktserniedrigung, weil die Siedelinie bzw. Soliduslinie für kleine Konzentrationen durch eine Gerade angenähert werden können. Am Siedepunkt der Lösung verdampft vorwiegend das Lösungsmittel, da der zu dieser Temperatur gehörige Taupunkt bei kleineren Konzentrationen liegt. Dies wird bei der Salzgewinnung durch Verdunsten des Meerwassers ausgenutzt.

In einer *echten Lösung* wird die feste Phase in ihre Atome oder Moleküle aufgetrennt. In einer Emulsion sind kleine Partikeln aufgeschwemmt. In *Elektrolyten* sind die Moleküle in Ionen aufgespalten worden. *Halbdurchlässige (semipermeable) Wände* sind poröse Schichten, die von bestimmten Molekülen durchdrungen werden können, von anderen nicht, z. B. in Abhängigkeit von der Molekülgröße. Ersetzt man die Trennwand durch eine semipermeable Schicht, die nur vom Lösungsmittel durchdrungen werden kann, dann tritt Lösungsmittel in die Lösung über, um diese zu verdünnen. Dem entgegen wirkt der entstehende Überdruck, der *osmotische Druck* p_{osm}, der sich im Gleichgewichtszustand ausbildet. Sein Wert $p_{osm} = z_L RT/V_{LM}$ ist gleich dem Druck, den die gelöste Substanz (z_L Mol) als Gas im Volumen V_{LM}, das das Lösungsmittel einnimmt, ausüben würde.

2.6.8.4. Mischentropie. Bei einer *idealen Mischung* zweier

Gase, Flüssigkeiten oder fester Phasen, die unter gleichem Druck stehen, ändert sich definitionsgemäß weder das Gesamtvolumen noch die innere Energie. Jedoch tritt eine Entropieänderung auf, weil das Mischen ein irreversibler Vorgang ist. Betrachtet man als Ausgangszustand zwei ideale Gase mit den Zustandsgrößen p, T und V_1, U_1, S_1 bzw. V_2U_2, S_2, die durch eine Wand getrennt sind, mischen sich die Gase, wenn die Trennwand entfernt wird. Druck und Temperatur bleiben unverändert, jedes Gas nimmt das Gesamtvolumen $V = V_1 + V_2$ ein. Da die innere Energie eines idealen Gases nicht vom Volumen abhängt, gilt $U = U_1 + U_2$ vor und nach dem Mischen.

Um die Entropieänderung zu bestimmen, muß man den gleichen Endzustand auf reversiblem Wege herstellen. Man erhält

$$\Delta S = R[(1-x)\ln(1-x) + x\ln x]$$

als Mischentropie je Mol. Sie verschwindet für die reinen Substanzen und hat bei $x = 0{,}5$ ihr Maximum.

2.6.8.5. Freie Enthalpie eines binären Systems. Das Minimum der freien Enthalpie bestimmt den thermodynamischen Gleichgewichtszustand bei gegebenen Werten der Zustandsgrößen p und T. Im binären System interessiert daher sowohl die Konzentrationsabhängigkeit der freien Enthalpie wie auch der Vergleich der freien Enthalpien verschiedener Phasen.

Beschränkt man sich auf die Betrachtung fester und flüssiger Phasen, dann kann man in der freien Enthalpie $G = U - TS + pV$ wegen der geringen Volumenänderungen die Konzentrationsabhängigkeit des Terms pV vernachlässigen, wenn man nicht das Verhalten unter hohem Druck untersucht. Die Konzentrationsabhängigkeit der Entropie kann durch die Mischentropie beschrieben werden. Die innere Energie ändert sich dadurch, daß A-A- bzw. B-B-Bindungen durch A-B-Bindungen ersetzt werden. Für die Betrachtungen ist es wesentlich, ob die Bindungsenergie je gleichartigem Paar (E_{AA}, E_{BB}) größer oder kleiner als die Bindungsenergie zwischen unterschiedlichen Partnern ist. Man geht davon aus, daß diese Verhältnisse einfach dadurch beschrieben werden können, daß je Mol A-B-Bindungen ein Energieaufwand $2E_{AB}$ erforderlich ist. Die Wahrscheinlichkeit, daß ein A-Atom auf einem Nachbarplatz ein B-Atom vorfindet, hängt in der Form $x(1-x)$ von der Konzentration ab.

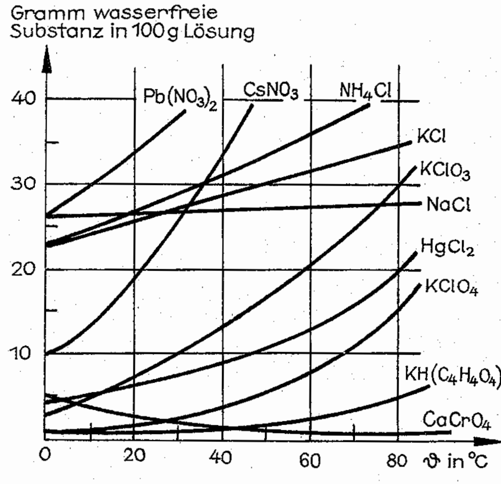


Abb. 2.6.8–8 Löslichkeit einiger Salze in Wasser (nach EGGERT)

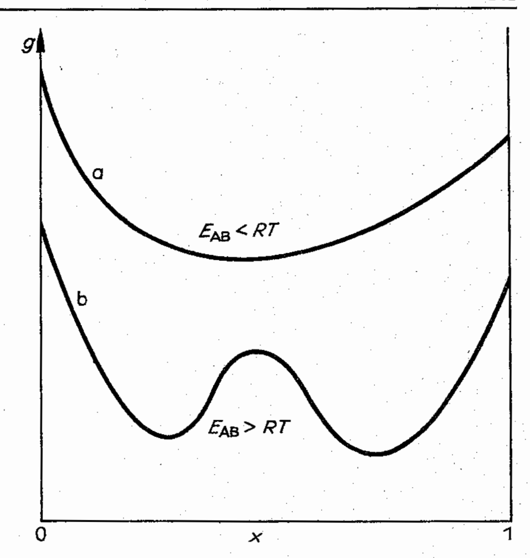


Abb. 2.6.8–9 Molare freie Enthalpie eines binären Systems

In diesem einfachen Bild ist die Konzentrationsabhängigkeit der molaren freien Enthalpie durch $g(x) = g(0) + 2E_{AB}x(1-x) + RT[(1-x)\ln(1-x) + x\ln x]$ gegeben (Abb. 2.6.8–9). Je nach dem, ob $E_{AB} > RT$ oder $E_{AB} < RT$ gilt, hat $g(x)$ ein Zwischenmaximum oder nicht. Der zweite Fall (b) ist für $E_{AB} < 0$ erfüllt, d. h., wenn die Bindung zwischen verschiedenartigen Atomen stärker ist als zwischen gleichartigen. Auch für $T \to \infty$, d. h. für die flüssige und gasförmige Phase, ist diese Bedingung immer erfüllt. Der erste Kurventyp ($E_{AB} > RT$) tritt nur auf, wenn sowohl die A-B-Bindungen schwächer sind als die zwischen gleichartigen Atomen ($E_{AB} > 0$) als auch die Temperatur niedrig genug ist. Aus diesen beiden Kurventypen kann man die Form von Zustandsdiagrammen (vgl. Abb. 2.6.8–2, 2.6.8–3) verständlich machen. Liegen bei einer Gesamtkonzentration x zwei Phasen (I, II) mit den Konzentrationen x^{I} und x^{II} vor, dann ist die molare freie Enthalpie $g(x) = \frac{x^{II}-x}{x^{II}-x^{I}} g^{I}(x^{I}) + \frac{x-x^{I}}{x^{II}-x^{I}} g^{II}(x^{II})$, wenn g^{I} und g^{II} die molaren freien Enthalpien der einzelnen Phasen sind. Für gegebenes x und gegebene Funktionen g^{I} und g^{II} sind die Konzentrationen x^{I} und x^{II} durch Minimalisierung von g zu finden.

Für die Schmelze (g^{I}) wird eine Kurve vom Typ (b) in Abb. 2.6.8–9 gewählt. Wird die feste Phase (g^{II}) durch den gleichen Kurventyp beschrieben, dann ergibt sich bei der in Abb. 2.6.8–10 dargestellten Temperaturabhängigkeit ein Zustandsdiagramm ohne Mischungslücke. Das Zweiphasengebiet zwischen Mischkristall und Schmelze ergibt sich dadurch, daß das Minimum von g in einem gewissen Temperaturbereich durch die gemeinsame Tangente von g^{I} und g^{II} gegeben ist. In den Grenzfällen, in denen im gesamten Konzentrationsbereich eine Kurve unter der anderen liegt, tritt die andere Phase gar nicht auf.

Abb. 2.6.8–11 zeigt, welches Zustandsdiagramm entsteht, wenn die molare freie Enthalpie des Kristalles durch eine Kurve vom Typ (a) der Abb. 2.6.8–9 beschrieben wird. Es entsteht eine Mischungslücke. Sie tritt also auf, wenn die Bindung zwischen gleichartigen Atomen stärker als zwischen ungleichartigen ist.

2.6.9. Substanzen unter dem Einfluß von Feldern

Auch über das Verhalten von Substanzen in elektrischen und magnetischen Feldern lassen sich aus den Hauptsätzen der Thermodynamik weitreichende Aussagen gewinnen. Elektrisches und magnetisches Feld stellen dabei zusätzli-

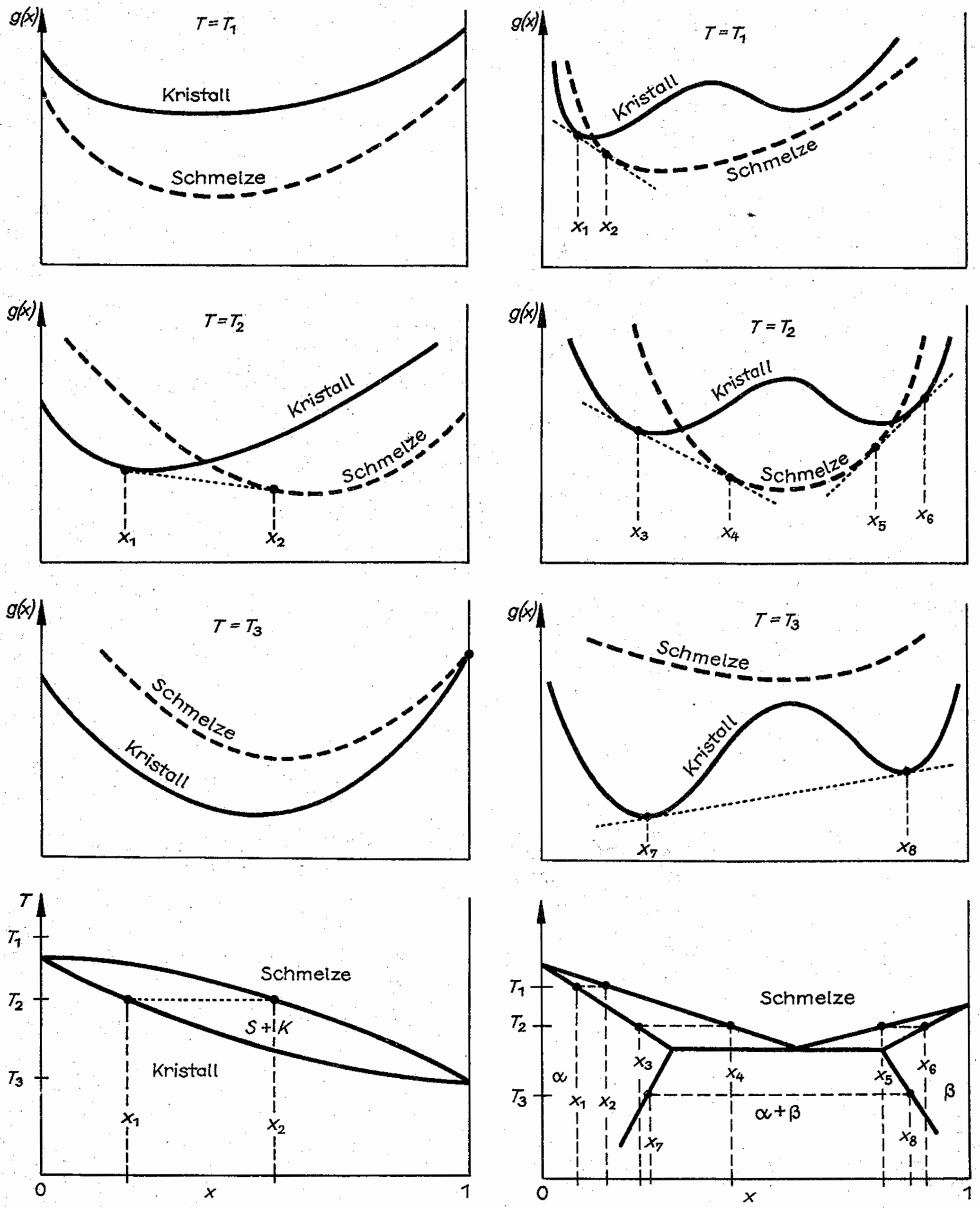


Abb. 2.6.8–10 Temperaturabhängigkeit der molaren freien Enthalpie bei lückenloser Mischbarkeit

Abb. 2.6.8–11 Temperaturabhängigkeit der molaren freien Enthalpie beim Auftreten einer Mischungslücke

che Freiheitsgrade dar, die zu den normalerweise zwei Freiheitsgraden (z. B. p, T) hinzukommen.
Die Kräfte auf ein Volumenelement im Festkörper lassen sich i. allg. nicht allein durch einen hydrostatischen Druck beschreiben, sie werden durch einen Spannungstensor charakterisiert. Anstelle der Variablen p und V treten dann Spannungs- und Verzerrungsfeld.

2.6.9.1. Thermodynamische Potentiale. Durch ein elektrisches Feld $\boldsymbol{E}$ entsteht in einer Substanz eine *Polarisation* $\boldsymbol{P}^*$ (elektrische Dipoldichte), durch ein Magnetfeld $\boldsymbol{H}$ eine *Magnetisierung* $\boldsymbol{M}^*$ (magnetische Dipoldichte). Dies geschieht durch Ausrichten vorhandener Dipole (Parelektrizität, Paramagnetismus), deren Richtung ohne äußeres Feld infolge der Wärmebewegung statistisch verteilt ist, oder (und) durch Induzieren von Dipolen (Dielektrizität, Diamagnetismus). Es wird eine auf ein Mol bezogene Größe eingeführt, das Dipolmoment je Mol $P = P^* V_m$ bzw. $M = M^* V_m$. Bei Anwesenheit eines elektrischen Feldes enthält die Gesamtenergie zunächst die im elektrischen Feld gespeicherte Energie. Zweitens hat ein Dipol die Wechselwirkungsenergie $W_D = -\boldsymbol{P} \cdot \boldsymbol{E}$. Schließlich ändert sich die Energie einer Substanz, wenn in ihr Dipole ausgerichtet oder induziert werden. Nur der letzte Anteil wird als Teil der inneren Energie gezählt, wobei der Zustand der Substanz durch den Wert der Polarisation (Magnetisierung) gekennzeichnet wird. Mit dieser Festlegung ist die Änderung der inneren Energie durch $\boldsymbol{E} \cdot \mathrm{d}\boldsymbol{P}$ gegeben. Analoges gilt im Magnetfeld. Bei Deformationen lautet – durch Spannungstensor σ_{ik} und Verzerrungstensor ε_{ik} ausgedrückt – die dem System zugeführte mechanische Arbeit $\sigma_{ik} \mathrm{d}\, \varepsilon_{ik}$. Bei hydrostatischem Druck $\sigma_{ik} = -p\, \delta_{ik}$ geht dieser Ausdruck wegen Spur $(\mathrm{d}\, \varepsilon_{kl}) = \mathrm{d}V$ in $-p\mathrm{d}V$ über.
Zusammenfassend ist die Änderung der inneren Energie bzw. der freien Energie bzw. der freien Enthalpie gegeben durch

$$\mathrm{d}U(S, \varepsilon, \boldsymbol{M}, \boldsymbol{P}) = T\mathrm{d}S + \sigma_{ik}\mathrm{d}\varepsilon_{kl} + \mu_0 \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{M} + \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{P},$$
$$\mathrm{d}F(T, \varepsilon, \boldsymbol{M}, \boldsymbol{P}) = -S\mathrm{d}T + \sigma_{ik}\mathrm{d}\varepsilon_{kl} + \mu_0 \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{M} + \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{P},$$
$$\mathrm{d}G(T, \sigma, \boldsymbol{M}, \boldsymbol{P}) = -S\mathrm{d}T - \varepsilon_{ik}\mathrm{d}\sigma_{kl} + \mu_0 \boldsymbol{H} \cdot \mathrm{d}\boldsymbol{M} + \boldsymbol{E} \cdot \mathrm{d}\boldsymbol{P}.$$

Auf Begriffsunterscheidungen, wie elastische oder elektrische Enthalpie, soll hier verzichtet werden. Unter der freien Enthalpie schlechthin dürfte man dann nur dasjenige thermodynamische Potential verstehen, das nur von intensiven Variablen abhängt, d. h. von T, σ, $\boldsymbol{H}$ und $\boldsymbol{E}$. Aus der oben eingeführten Funktion G berechnen sich die Felder, die die betrachtete Polarisation (Magnetisierung) bewirken, zu $\mu_0 \mathrm{H} = (\partial G/\partial \boldsymbol{M})_{T,\sigma}$, $\boldsymbol{E} = (\partial G/\partial \boldsymbol{P})_{T,\sigma}$. Mit diesen erweiterten thermodynamischen Potentialen lassen sich Erscheinungen wie Ferromagnetismus $\boldsymbol{M}(\boldsymbol{H}, T)$, Ferroelektrizität $\boldsymbol{P}(\boldsymbol{E}, T)$, Thermoelastizität $\varepsilon(T)$, Magnetostriktion $\varepsilon(\boldsymbol{M})$, Elektrostriktion $\varepsilon(\boldsymbol{E})$ und Piezoelektrizität $\boldsymbol{P}(\sigma)$ diskutieren.

2.6.9.2. Phasenumwandlung mit Ordnungsparameter, Ferromagnetismus. Am Phasenübergang zwischen paramagnetischem und ferromagnetischem Zustand ist die spontane Magnetisierung, d. h. die Magnetisierung ohne äußeres Magnetfeld, Null. Unterhalb der Umwandlungstemperatur T_c, der *Curie-Temperatur*, wächst sie im ferromagnetischen Zustand von Null beginnend bei weiterer Abkühlung an (Abb. 2.6.9–1b). Eine Größe, die sich in dieser Weise verhält, nennt man einen Ordnungsparameter: Oberhalb T_c, in der ungeordneten Phase, hat er den Wert Null, unterhalb von T_c, in der geordneten Phase, wächst er von Null beginnend an. Im betrachteten Fall ist die Magnetisierung der

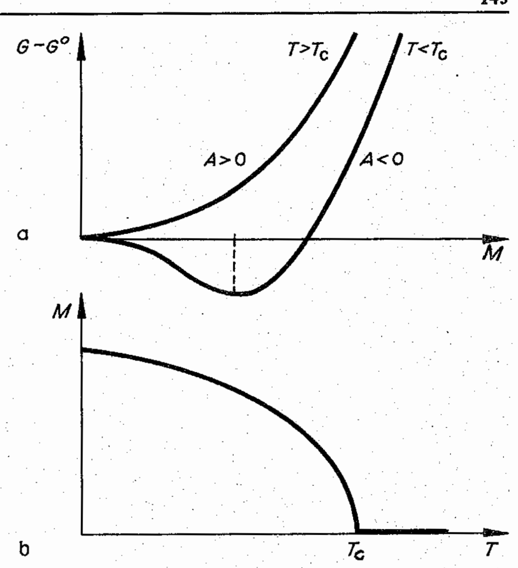


Abb. 2.6.9–1 (a) Verlauf der freien Enthalpie und (b) Temperaturabhängigkeit der spontanen Magnetisierung bei einem Phasenübergang zweiter Ordnung

Ordnungsparameter, oft bedarf es detaillierter Betrachtungen, um den Ordnungsparameter einer Phasenumwandlung zu finden.
Die *Landau-Theorie der Phasenumwandlungen* geht davon aus, daß die thermodynamischen Potentiale in der Umgebung der Umwandlungstemperatur nach Potenzen des Ordnungsparameters entwickelbar sind. Das magnetische Verhalten eines homogenen, isotropen Materiales wird also durch die freie Enthalpie $G(p, T, M) = G(p, T, 0) + AM^2 + BM^4 + CM^6 + \ldots$ beschrieben. Ohne äußeres Magnetfeld muß $(\partial G/\partial M)_{T,p} = 0$ gelten, im thermodynamischen Gleichgewicht hat G ein Minimum.
A, B und C sind Funktionen von Druck und Temperatur. Im weiteren wird nur die Abhängigkeit von der Temperatur betrachtet. Zunächst wird angenommen, daß im interessierenden Temperaturbereich $B > 0$ und $C > 0$ gilt. Dann kann wegen der kleinen M-Werte der Summand CM^6 vernachlässigt werden. Der Koeffizient A soll beide Vorzeichen annehmen können, bei einer Temperatur T_c soll er sein Vorzeichen wechseln $A = a(T - T_c)$, $a > 0$. Die freie Enthalpie als Funktion der Magnetisierung hat dann den in Abb. 2.6.9–1a dargestellten Verlauf. Sie hat ein Minimum für den Wert der Magnetisierung $M = [(a/2B)(T_c - T)]^{1/2}$ für $T < T_c$ bzw. $M = 0$ für $T > T_c$. Unterhalb von T_c tritt, wie in Abb. 2.6.9–1b gezeigt wird, eine spontane Magnetisierung auf. Für $T \to T_c$ geht M stetig in den Wert Null über, es liegt eine *Phasenumwandlung zweiter Ordnung* vor.
Mit $A = a(T - T_c)$, $B < 0$ und $C > 0$ hat die freie Enthalpie den in Abb. 2.6.9–2a gezeigten Verlauf. Die aus der Minimalisierungsforderung folgende Magnetisierung ist in Abb. 2.6.9–2b dargestellt.

$$M = \left\{ \frac{|B|}{3C} \left[1 + \left(1 - \frac{3}{4} \frac{T - T_c}{T_0 - T_c} \right)^{1/2} \right] \right\}^{1/2},$$

$$T_0 = T_c + \frac{B^2}{4aC}.$$

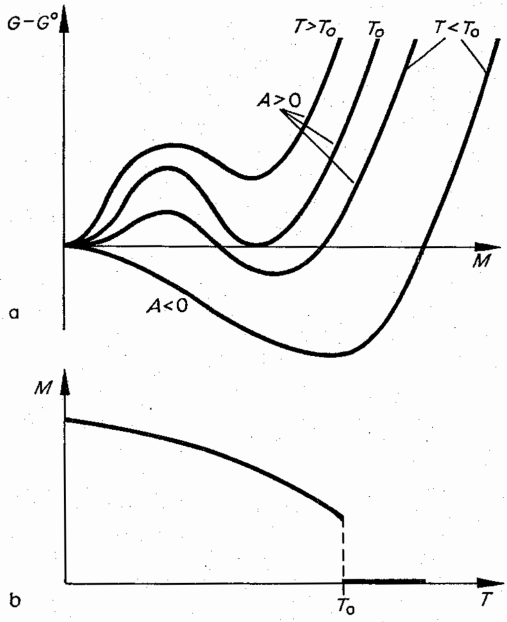


Abb. 2.6.9–2 (a) Verlauf der freien Enthalpie und (b) Temperaturabhängigkeit der spontanen Magnetisierung bei einem Phasenübergang erster Ordnung

Bei der Temperatur $T_0 (> T_c)$ stehen zwei Phasen miteinander im Gleichgewicht, eine mit $M = 0$ und eine mit einem endlichen Wert der Magnetisierung. Auch diese *Phasenumwandlung erster Ordnung* kann mit der Landau-Theorie der Phasenumwandlungen beschrieben werden.

2.6.9.3. Ferroelektrizität. Ferroelektrizität wird mit der Landau-Theorie analog zum Ferromagnetismus beschrieben. Die Betrachtungen sollen hier auf einen kubischen Kristall erweitert werden, um die erforderlichen Beziehungen bei einem *anisotropen System* zu erläutern. Es tritt nicht nur der Betrag der Polarisation auf, sondern auch die einzelnen Komponenten dieses Vektors, jedoch nur in solchen Kombinationen, die die kubische Symmetrie widerspiegeln. Wegen des Inversionszentrums treten nur gerade Potenzen auf:

$$G(p, T, \boldsymbol{P}) = G(p, T, 0) + a(T - T_c)(P_x^2 + P_y^2 + P_z^2)$$
$$+ B_1(P_x^4 + P_y^4 + P_z^4) + 2B_2(P_x^2P_y^2 + P_y^2P_z^2 + P_z^2P_x^2).$$

Es wird der auf eine Phasenumwandlung zweiter Ordnung führende Fall $B_1, B_2 > 0$ betrachtet. Für $B_1 = B_2$ erhält man wieder den vom isotropen Medium her bekannten Ausdruck. Die Lösungen

$$P_x = P_y = P_z = 0 \qquad \text{für } T > T_c,$$

$$P_x = P, \quad P_y = P_z = 0, \quad P = \sqrt{\frac{a(T_c - T)}{2B_1}} \qquad \text{für } T < T_c,$$

$$P_x = P_y = P_z = P, \qquad P = \sqrt{\frac{a(T_c - T)}{B_1 + B_2}} \qquad \text{für } T < T_c$$

folgen aus der Minimalisierungsforderung, wobei die Lösung 2 für $B_1 < B_2$, die Lösung 3 für $B_1 > B_2$ stabil ist.

Es zeigt sich, daß in einem kubischen Kristall eine spontane Polarisation nur in [100]- oder [111]-Richtung (oder kristallografisch äquivalenten Richtungen) existieren kann. Da die Koeffizienten B_1, B_2 von der Temperatur abhängen, ist ein Umschlagen der Lösungen 2 und 3 in Abhängigkeit von T denkbar.

2.6.9.4. Hysterese. Die Betrachtungen laufen für ferromagnetische und ferroelektrische Substanzen ganz analog. Im homogenen System, charakterisiert durch $G(T, p, M)$, ergibt sich der Zusammenhang zwischen Magnetisierung und Feld aus $\mu_0 H = (\partial G/\partial M)_{T,p}$. Mit $G(T, p, M) = G(T, p, 0) + a(T - T_c) M^2 + BM^4$ $(B > 0)$ ist das die in Abb. 2.6.9–3 dargestellte Kurve dritten Grades $\mu_0 H = 2a(T - T_c)\mathrm{H} + 4\mathrm{BM}^3$. Für $T > T_c$ ergibt sich für kleine Felder ein linearer Zusammenhang zwischen M und H mit der als *Curie-Weiß-Gesetz* bezeichneten Temperaturabhängigkeit der magnetischen Suszeptibilität $\chi = \partial M/\partial H = C/(T - T_c)$, wie sie für paramagnetische Substanzen charakteristisch ist. Die auftretenden Faktoren wurden in C zusammengefaßt. Für $T < T_c$ gibt es für kleine H drei Lösungen M. Eine Untersuchung der Stabilität dieser Lösungen zeigt, daß der durch Pfeile gekennzeichnete Weg durchlaufen wird. Damit wird die Hysterese aus diesen Betrachtungen gewonnen. Das scharfe Umschlagen der Magnetisierung vom Punkt 4 nach 1 bzw. von Punkt 2 nach 3 wird experimentell nicht beobachtet (gepunktete Kurve in Abb. 2.6.9–3b), weil einzelne Bezirke der Substanz ihre Magnetisierung nacheinander wechseln.

2.6.9.5. Thermoelastizität. Die Hauptsätze der Thermodynamik führen auf $\mathrm{d}F = S\mathrm{d}T + \mathrm{Spur}(\sigma_{ik}\mathrm{d}\varepsilon_{ki})$, woraus σ_{ik}

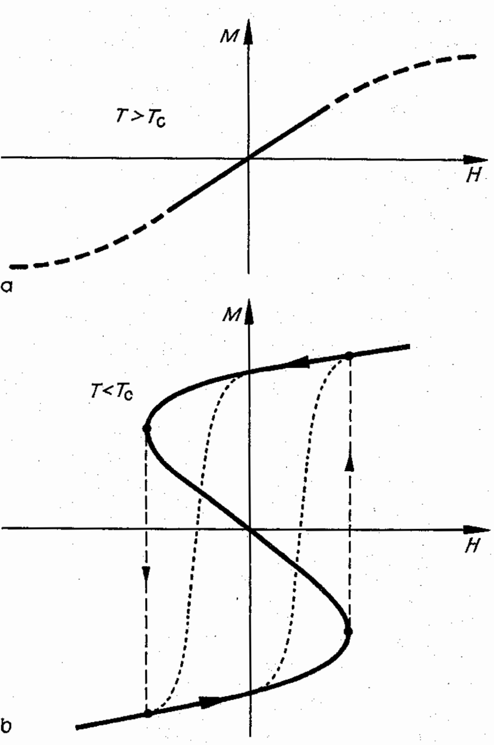


Abb. 2.6.9–3 Feldstärkeabhängigkeit der Magnetisierung, (a) paramagnetischer Fall, (b) Hysterese beim Ferromagnetismus

$= (\partial F/\partial \varepsilon_{ik})_T$ folgt. Bei kleinen Verzerrungen hat die freie Energie die Form $F(T, \varepsilon) = F(T, 0) - \left(\lambda + \frac{2}{3}\varkappa\right)\alpha(T - T_0)$ $\mathrm{Spur}(\varepsilon) + \frac{\lambda}{2}(\mathrm{Spur}\,\varepsilon)^2 + \varkappa\ \mathrm{Spur}(\varepsilon^2)$, wobei bei der Temperatur $T(\boldsymbol{r}) = T_0 =$ konst. keine Wärmespannungen auftreten sollen. Daraus folgt für den Spannungstensor $\sigma = -\left(\lambda + \frac{2}{3}\varkappa\right)\alpha(T - T_0)\mathrm{I} + 2\varkappa\varepsilon + \lambda \mathrm{I}\ \mathrm{Spur}\,\varepsilon$ (I = Einheitstensor). Für $T = T_0$ vereinfacht sich diese Beziehung zu $\sigma = 2\varkappa\varepsilon + \lambda\ \mathrm{I}\ \mathrm{Spur}\,\varepsilon$. Das ist das Hookesche Gesetz für ein homogenes System. $\varkappa$ und λ heißen *Lamésche Parameter*, sie können durch Elastizitätsmodul E und Poisson-Konstante ausgedrückt werden (vgl. 2.2.). Aus dem spannungsfreien Fall erkennt man mit Spur $\varepsilon = \Delta V/V$, daß α der Koeffizient der Wärmeausdehnung ist.

Für den Verschiebungsvektor $\boldsymbol{s}(\boldsymbol{r})$ erhält man mit $2\varepsilon_{ik} = \partial s_i/\partial x_k + \partial s_k/\partial x_i$ im statischen Fall ($\nabla\sigma = 0$) die Bestimmungsgleichung $\Delta \boldsymbol{s} + (1 - 2\nu)^{-1}\nabla(\nabla \cdot \boldsymbol{s}) = (2/3)(1 + \nu)(1 - 2\nu)^{-1}\alpha\nabla T$, aus der bei gegebenem Temperaturverlauf $T(\boldsymbol{r})$ eine Lösung so zu bestimmen ist, daß die für $\boldsymbol{s}$, ε oder σ geforderten Randbedingungen erfüllt werden. Aus dieser Lösung kann dann der Verzerrungstensor ε bzw. über das Hookesche Gesetz der Spannungstensor σ berechnet werden. Diese Formeln stellen das mathematische Gerüst dar, um die Spannungen zu berechnen, die beim ungleichmäßigen Erwärmen eines Körpers auftreten. Wie das Zerspringen einer Glasschale beim Eingießen einer kochenden Flüssigkeit zeigt, können diese Spannungen so groß sein, daß sie zur Zerstörung des Körpers führen.

2.6.10. Hohlraumstrahlung

Hier werden Fragen der Thermodynamik der Hohlraumstrahlung besprochen. Ihre Rolle bei der Entwicklung der Quantenphysik wird in Kapitel 4 behandelt.

2.6.10.1. Spektrales Emissionsvermögen. Die Substanzen sind aus geladenen Teilchen aufgebaut. Wenn sich diese Ladungen beschleunigt bewegen, wie sie es infolge der Wärmebewegung tun, dann strahlen sie elektromagnetische Wellen ab. Es entsteht die Wärmestrahlung. Intensität und Frequenzverteilung der Strahlung hängen von der Temperatur ab. Mit steigender Temperatur wächst die Intensität, und das Maximum der Intensität verschiebt sich zu höheren Frequenzen bzw. kürzeren Wellenlängen hin. Die von einem Körper abgestrahlten elektromagnetischen Wellen charakterisiert man daher durch eine von Frequenz und Temperatur abhängige Funktion, das *spektrale Emissionsvermögen* $u(\nu, T)$.

Die auf einen Körper auffallende Strahlung kann reflektiert, absorbiert oder durchgelassen werden. Ist $I(\nu)\,\mathrm{d}\nu$ die Intensität der Strahlung im Frequenzbereich $\nu \dots \nu + \mathrm{d}\nu$, dann lautet diese Bilanz $I(\nu)\,\mathrm{d}\nu = I_\mathrm{R}(\nu)\,\mathrm{d}\nu + I_\mathrm{A}(\nu)\,\mathrm{d}\nu + I_\mathrm{D}(\nu)\,\mathrm{d}\nu$. Davon ausgehend lassen sich *Reflexionsvermögen* $R(\nu) = I_\mathrm{R}(\nu)/I(\nu)$ *Absorptionsvermögen* $A(\nu) = I_\mathrm{A}(\nu)/I(\nu)$ und *Durchlässigkeit* $D(\nu) = I_\mathrm{D}(\nu)/I(\nu)$ einführen. Es gilt dann $\boldsymbol{R}(\nu) + \boldsymbol{A}(\nu) + \boldsymbol{D}(\nu) = 1$. Einen Körper, der die gesamte auf ihn fallende Strahlung absorbiert ($A = 1$), nennt man einen *schwarzen Körper*, einen Körper mit $R = 1$ einen *weißen* oder *ideal spiegelnden Körper* und einen Körper mit $D = 1$ einen *absolut durchlässigen Körper*.

Eine besondere Rolle bei der Untersuchung der Strahlungsgesetze spielt der schwarze Körper. Die elektromagnetische Strahlung, die sich im thermodynamischen Gleichgewicht in einem von einem schwarzen Körper umgebenen Hohlraum befindet, nennt man *Hohlraumstrahlung*. Wegen des hohen Absorptionsvermögens $A = 0{,}95$ von Ruß kann man einen schwarzen Körper experimentell näherungsweise realisieren, indem man die Innenwände eines Behälters mit Ruß bestreicht. Ein kleines Loch im Behälter beeinflußt die Strahlung im Hohlraum nur unwesentlich. Die aus dem Loch austretende Strahlung hat daher die Temperatur- und Frequenzabhängigkeit der Hohlraumstrahlung. Die Hohlraumstrahlung ist homogen, isotrop und unpolarisiert. Sie läßt sich durch die *spektrale Energiedichte* $\eta(\nu, T)$ charakterisieren. Die Energiedichte $\eta(T)$ ergibt sich daraus durch Integration über alle Frequenzen. $\eta(\nu, T)$ hängt nicht von der Größe und Form des Hohlraumes ab. Die innere Energie U der Strahlung ist daher zum Volumen V proportional

$$U(V, T) = V\eta(T) = V\int_0^\infty \mathrm{d}\nu\,\eta(\nu, T).$$

Da ein schwarzer Körper die gesamte auf ihn treffende Strahlung absorbiert, muß er in einer solchen Weise Strahlung emittieren, daß das thermodynamische Gleichgewicht, d. h. die spektrale Energiedichte der Hohlraumstrahlung, nicht gestört wird. Daraus folgt der Zusammenhang $u(\nu, T) = (c/4)\,\eta(\nu, T)$ zwischen spektraler Energiedichte und spektralem Emissionsvermögen. Wird ein nichtschwarzer Körper in den Hohlraum gebracht, dann kann das thermodynamische Gleichgewicht nicht geändert werden. Das heißt, die auf ihn treffende Strahlung wird durch $\eta(\nu, T)$ bzw. durch das Emissionsvermögen $u(\nu, T)$ des schwarzen Strahlers bestimmt. Die Bilanz von Absorption und Emission führt auf den *Kirchhoffschen Satz*: Das Verhältnis von spektralem Emissionsvermögen $\tilde{u}(\nu, T)$ und spektralem Absorptionsvermögen $A(\nu)$ ist bei gegebener Temperatur von den spezifischen Eigenschaften eines Körpers unabhängig, d. h. für alle Körper gleich und gleich dem spektralen Emissionsvermögen des schwarzen Körpers $\tilde{u}(\nu, T)/A(\nu) = u(\nu, T)$. Frequenz- und Temperaturabhängigkeit dieser universellen Funktion

$$u(\nu, T) = 2\pi\frac{h\nu^3}{c^2}\frac{1}{\mathrm{e}^{h\nu/(kT)} - 1}, \quad \int_0^\infty \mathrm{d}\nu u(\nu, T) = u(T),$$

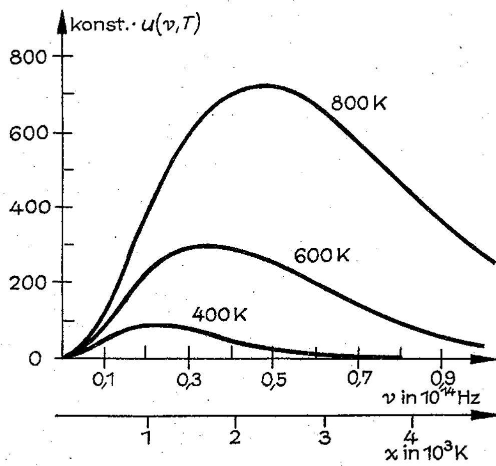


Abb. 2.6.10–1 Spektrales Emissionsvermögen des schwarzen Körpers, konst. $\cdot\, u = x^3/(\mathrm{e}^{x/(T)} - 1)$, $x = h\nu/k$

$$u(\lambda, T) = 2\pi \frac{hc^2}{\lambda^5} \frac{1}{e^{hc/(kT\lambda)} - 1}, \quad \int_0^\infty d\lambda\, u(\lambda, T) = u(T)$$

sind durch die Plancksche Strahlungsformel (Abb. 2.6.10–1) gegeben. Da eine Fläche in den Raumwinkel 2π abstrahlt, ergibt sich aus dieser Formel das spektrale Emissionsvermögen je Raumwinkel u_Ω zu $u_\Omega(\nu, T) = u(\nu, T)/2\pi$ (k = Boltzmann-Konstante, h = Plancksches Wirkungsquantum).

2.6.10.2. Strahlungsdruck. Elektromagnetische Strahlung transportiert nicht nur Energie, sondern auch Impuls. Die *Impulsdichte* $\pi = \eta/c$ ist durch die Energiedichte bestimmt. Dieser Zusammenhang entspricht dem Zusammenhang $E = cP$ zwischen Energie und Impuls der Photonen. Bei der Absorption der Strahlung wird deren Impuls von der Wand aufgenommen, bei der Emission wird auf die Wand ein Rückstoßimpuls übertragen. Dieser Impuls ist proportional zur Energiedichte, zur Fläche und zum betrachteten Zeitintervall. Die auf die Wand wirkende Kraft ist gleich dem je Zeiteinheit übertragenen Impuls, und der Druck ergibt sich als Kraft je Fläche. Damit wird der *Strahlungsdruck* proportional zur Energiedichte $p = \eta/3$. Den Zahlenfaktor erhält man aus der Winkelintegration.

2.6.10.3. Stefan-Boltzmann-Gesetz. Die Gleichung für den Strahlungsdruck $p(V, T) = \eta(T)/3$ stellt die thermische Zustandsgleichung für die Hohlraumstrahlung dar. Der Druck ist unabhängig vom Volumen. In der kalorischen Zustandsgleichung $U(V, T) = V\eta(T)$ steht auch nur die Energiedichte als zunächst unbekannte Funktion der Temperatur, die Volumenabhängigkeit ist bekannt.
Die aus dem zweiten Hauptsatz der Wärmelehre folgende Beziehung $(\partial U/\partial V)_T = T(\partial p/\partial T)_V - p$ zwischen der Volumenabhängigkeit der inneren Energie und der thermischen Zustandsgleichung kann daher genutzt werden, um die Funktion $\eta(T)$ zu bestimmen.

$$\eta(T) = aT^4, \quad u(T) = \sigma T^4$$

wird als Stefan-Boltzmannsches Gesetz bezeichnet. Die Zahlenwerte $\sigma = k^4\pi^2/(60c^2\hbar^3) = 5{,}670 \cdot 10^{-8}\,\mathrm{W/(m^2K^4)}$ bzw. $a = \sigma/c$ ergeben sich experimentell bzw. mit Hilfe der Planckschen Strahlungsformel. Aus der Entropie als thermodynamischen Potential

$$S(U, V) = (4/3)\, a^{1/4}\, U^{3/4}\, V^{1/4} = (4/3)\, a\, T^3\, V$$

lassen sich alle thermodynamischen Größen berechnen.

2.6.10.4. Wiensches Gesetz. Die Hohlraumstrahlung wird durch eine Funktion, das spektrale Emissionsvermögen $u(\nu, T)$, charakterisiert. Das Stefan-Boltzmannsche Gesetz ist eine Aussage über diese Funktion, die aus rein thermodynamischen Überlegungen gewonnen wurde. WIEN verfeinerte diese Aussagen. In einem Gedankenexperiment betrachtete er die adiabatische Kompression von Hohlraumstrahlung in einem Zylinder mit spiegelnden Wänden. Bei einem adiabatischen Prozeß bleibt die Entropie konstant. Bei der Kompression wird Arbeit gegen den Strahlungsdruck geleistet. Innere Energie und Temperatur erhöhen sich. Bei der Reflexion an dem bewegten Kolben wird die Frequenz der elektromagnetischen Wellen verändert. Bei quasistatischer Führung der adiabatischen Kompression, d. h. wenn ständig Gleichgewichtszustände durchlaufen werden, muß trotz Frequenzänderung und Temperaturerhöhung die spektrale Energiedichte durch die Funktion $\eta(\nu, T)$ gegeben sein. Daraus folgt für η bzw. für u das Wiensche Gesetz $u(\nu, T) = \nu^3 f(\nu/T) = T^3 \tilde{f}(\nu/T)$.
Es verbleibt nur noch eine Funktion von einer Variablen zu bestimmen. Über diese können aus rein thermodynamischen Überlegungen keine Aussagen gemacht werden. Experimentell kann ihre Form gemessen werden (vgl. Abb. 2.6.10–1). Ihre Berechnung, die Aussagen der Quantenphysik erfordert, führt auf die Plancksche Strahlungsformel. Trägt man das spektrale Emissionsvermögen $u(\nu, T)$ über ν/T auf, dann wächst die Amplitude mit T^3, die Form der Kurve bleibt unverändert. Insbesondere folgt daraus ν_{max}/T = konst. bzw.

$$\lambda_{max}\, T = \text{konst.} = 2{,}898 \cdot 10^{-3}\,\mathrm{m \cdot K},$$

das sogenannte *Wiensche Verschiebungsgesetz.*

2.6.11. Irreversible Prozesse

Die reversible, quasistatisch ablaufende Zustandsänderung ist eine Idealisierung, die nicht realisiert werden kann. Es treten immer endliche Temperatur- oder Druckunterschiede auf. Damit verbunden sind Ströme, Transportvorgänge. Die Zustandsänderungen verlaufen irreversibel. Auf die Transporterscheinungen soll hier eingegangen werden.

2.6.11.1. Lokales thermodynamisches Gleichgewicht. Thermodynamische Betrachtungen sind auch auf Zustände nahe dem Gleichgewichtszustand sehr erfolgreich anwendbar, insbesondere, wenn sich eine *lokale Temperatur* angeben läßt. Das betrachtete System wird in Volumenelemente unterteilt (Abb. 2.6.11–1). Jedes Volumenelement wird für sich als ein thermodynamisches System konstanter Temperatur betrachtet, das in Kontakt mit anderen Systemen steht, deren Temperatur infinitesimal abweicht. Die Temperaturverteilung der einzelnen Systeme – Volumenelemente in der Umgebung verschiedener Orte $\boldsymbol{r}$ – wird durch ein *Temperaturfeld* $T(\boldsymbol{r}, t)$ beschrieben. Voraussetzung dafür, daß eine solche Betrachtung sinnvoll ist, ist eine schwache Ortsabhängigkeit. Die Volumenelemente müssen so groß wählbar sein, daß sich in ihnen noch viele Teilchen (Atome oder Moleküle) befinden und daß die mittlere freie Weglänge der Teilchen klein gegen die Ausdehnung des Volumenelementes ist.
Thermodynamische Zustandsgrößen werden durch Felder ersetzt, die unter dieser Voraussetzung nur über die Tempe-

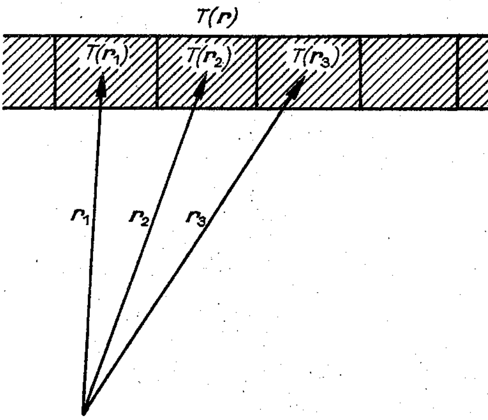


Abb. 2.6.11–1 Temperatur in verschiedenen Volumenelementen

ratur $T(r, t)$ oder die Dichte $\varrho(r, t)$ von Ort und Zeit abhängen. Die Felder sind die auf das Volumenelement (oder auch Massenelement) bezogenen Zustandsgrößen. Über $U = \int \mathrm{d}\tau u$, $S = \int \mathrm{d}\tau s$, $M = \int \mathrm{d}\tau \varrho$ usw. werden die Dichte der inneren Energie u, die Entropiedichte s oder die Massendichte ϱ definiert. Erster und zweiter Hauptsatz der Thermodynamik gelten für diese Größen in der bekannten Form.

2.6.11.2. Thermodynamische Kräfte und Ströme. Irreversible Prozesse sind durch das Auftreten von Strömen gekennzeichnet. Bei einem Konzentrationsgefälle tritt ein *Massenstrom* j_m auf, bei einem Temperaturgefälle ein *Wärmestrom* j_w und in einem elektrischen Feld ein *elektrischer Strom* j. Die Ursachen der Ströme nennt man thermodynamische Kräfte X_i. Der durch die Ströme bewirkte *Entropiezuwachs* ist durch $\dot{s} = \sum_i j_i X_i$ gegeben.

Über die Abhängigkeit der Stromdichten von den Kräften $j_i(X_k)$ lassen sich phänomenologische Aussagen machen. Bei verschwindenden Kräften treten keine Ströme auf. Dieser Zustand ist der Gleichgewichtszustand. Wird das Vorzeichen der Kräfte umgekehrt, dann drehen auch die Ströme ihr Vorzeichen um. So fließt der Wärmestrom immer vom wärmeren zum kälteren Wärmespeicher. Weiterhin kann man bei kleinen Abweichungen vom Gleichgewichtszustand, bei schwachen Kräften, Terme in höheren Potenzen der Kräfte vernachlässigen. Auf eine äußere Störung (thermodynamische Kraft) folgt also eine *lineare Reaktion*, die Stromdichten sind proportional zu den Kräften

$$j_i = \sum_k L_{ik} X_k .$$

Die *Transportkoeffizienten* L_{ik} sind in der Thermodynamik phänomenologische Größen, die experimentell bestimmt werden müssen. Der zweite Hauptsatz liefert über sie nur Aussagen von der Art, daß die Wärmeleitfähigkeit positiv ist. Die Berechnung durch mikroskopische Betrachtungen erfordert für jede Erscheinung gesonderte Überlegungen. Als allgemeine Aussage ist nur die *Onsagersche Reziprokitätsbeziehung* $L_{ik} = L_{ki}$ zu nennen, die aus der Invarianz der mikroskopischen Bewegungsgleichungen gegen Zeitumkehr folgt.

2.6.11.3. Wärmeleitung. Wärme kann durch Leitung, Konvektion oder Strahlung übertragen werden. Die *abgestrahlte Energie* wächst mit der vierten Potenz der Temperatur (vgl. 2.6.10.) und spielt meist nur bei hohen Temperaturen eine Rolle. Als *Konvektion* bezeichnet man die mit Massentransport verbundene Wärmeübertragung, die Fernheizung ist ein Beispiel dafür. In Gasen und Flüssigkeiten treten Wärmeleitung und -konvektion oft gekoppelt auf, in Festkörpern kann man reine *Wärmeleitung* am leichtesten beobachten.

In Abb. 2.6.11–2 sind zwei Wärmebehälter $T_1 > T_2$ durch einen Stab miteinander gekoppelt. Der Mantel des Stabes sei von einer wärmeisolierenden Schicht umgeben, so daß die Wärmeabgabe an die Umgebung vernachlässigt werden kann. Der Wärmestrom I_w, d. h. die je Zeit vom Behälter 1 auf den Behälter 2 übertragene Wärmemenge, ist proportional zum Temperaturunterschied $T_1 - T_2$. Der Temperaturunterschied stellt die treibende Kraft für den Wärmestrom dar. Der Wärmestrom ist der Länge des Stabes umgekehrt proportional. Dies kann man aus den in Abb. 2.6.11–3 dargestellten Versuchsbedingungen ersehen. Im ersten Fall überbrücken beide Stäbe einen Temperaturunterschied ΔT,

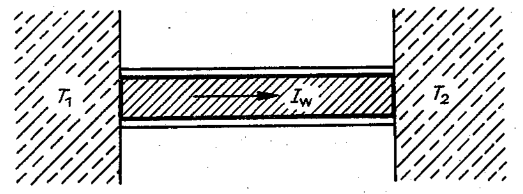


Abb. 2.6.11–2 Wärmestrom zwischen zwei Wärmebehältern

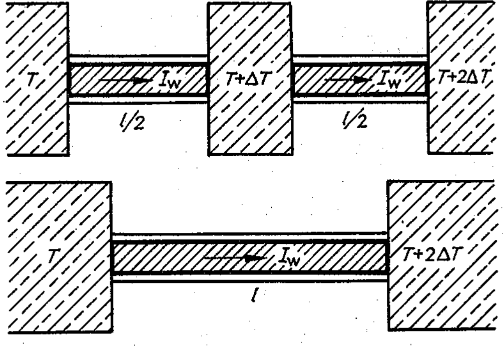


Abb. 2.6.11–3 Zur Abhängigkeit des Wärmestromes von der Länge des Stabes

im zweiten Fall ein Stab doppelter Länge einen Temperaturunterschied $2\Delta T$. In beiden Fällen ist der Wärmestrom gleich groß. Weiterhin ist der Wärmestrom proportional zur Querschnittsfläche des Stabes. Eine Verdopplung der Querschnittsfläche ist dem Anbringen eines zweiten Stabes äquivalent. Der Wärmestrom ist also durch $I_w = \varkappa(T_1 - T_2)(l/A) = (T_1 - T_2)/R$ gegeben. Darin ist $\varkappa$ die als *Wärmeleitfähigkeit* bezeichnete Materialkonstante (Tab. 2.6.11–4). In der zweiten Schreibweise hat die Gleichung für den Wärmestrom eine zum Ohmschen Gesetz analoge Form. Statt der Spannungsdifferenz steht die Temperaturdifferenz. R wird daher als *Wärmewiderstand* bezeichnet. Die Wärmewiderstände verschiedener Schichten kann man wie in Reihe geschaltete elektrische Widerstände addieren.

Gase haben eine kleine Wärmeleitfähigkeit. Deswegen kann man mit einer Luftschicht eine gute Wärmeisolation erreichen. Dies wird beim Doppelfenster, beim Federbett

Tab. 2.6.11-4 Wärmeleitfähigkeit bei Normaldruck

Substanz[1]	$\varkappa$ in W/m · K	Substanz[2]	$\varkappa$ in W/m · K
Ag	422	Si	126
Al	226	W	168
Au	298	He	0,143
Cu	402	Ar	0,0164
Fe	73,3	H_2	0,171
Li	71,2	O_2	0,0245
Ni	58,6	N_2	0,0240
Pb	34,8	Luft	0,0241
Pt	73	H_2O[3]	0,598

[1] bei 20 °C nach Schulze, [2] bei 0 °C nach Landolt-Börnstein, [3] bei 20 °C und Sättigungsdruck

oder bei Pelzbekleidung ausgenutzt. Auch Wasser ist ein schlechter Wärmeleiter. Man kann in einem Reagenzglas Wasser zum Kochen bringen, bevor Eis am Boden des Glases geschmolzen ist. Gute elektrische Leiter, wie Kupfer, Silber, haben auch eine hohe Wärmeleitfähigkeit. Das ist dadurch zu erklären, daß in Metallen die Wärmeleitung im wesentlichen über die Leitungselektronen erfolgt. Nach dem *Wiedemann-Franzschen Gesetz* $\varkappa/(\sigma T) = \text{konst.}$ sind elektrische Leitfähigkeit und Wärmeleitfähigkeit von Metallen zueinander proportional.

In dem einfachen Fall der Abb. 2.6.11–1 ist der Temperaturverlauf als linearer Abfall von T_1 nach T_2 vorgegeben. Im allgemeinen ist das Temperaturfeld $T(\boldsymbol{r}, t)$ zeitabhängig und in komplizierterer Weise vom Ort abhängig. Zu seiner Berechnung müssen drei Aussagen herangezogen werden: 1. die Bilanzgleichung für die innere Energie $\partial u/\partial t + \nabla \cdot \boldsymbol{j}_w + \nu = 0$; 2. die Materialgleichung $\boldsymbol{j}_w = -\varkappa \nabla T$, die den Temperaturgradienten als Ursache der Wärmestromdichte ausweist und die Wärmeleitfähigkeit als Materialkonstante enthält; 3. die Hauptsätze der Thermodynamik. Wird hier von Wärmeausdehnung und -spannungen abgesehen, dann reduzieren sie sich auf die Aussagen $\mathrm{d}u = T\mathrm{d}s$. Energiedichte u bzw. Entropiedichte s hängen nur über $T(\boldsymbol{r}, t)$ von der Zeit ab. Daher wird $\partial u/\partial t = c\varrho(\partial T/\partial t)$, und für die Temperaturverteilung ergibt sich die *Wärmeleitungsgleichung*

$$c\varrho(\partial T/\partial t) - \varkappa \Delta T + \nu = 0$$

als Bestimmungsgleichung für $T(\boldsymbol{r}, t)$. Über den ersten Hauptsatz konnte die spezifische Wärme c eingeführt werden, der zweite Hauptsatz verlangt $\varkappa > 0$ (ϱ = Massendichte, ν = Verlustleistungsdichte). Für den wichtigen Fall, daß durch einen elektrischen Strom Wärme erzeugt wird, gilt $\nu = -\boldsymbol{j} \cdot \boldsymbol{E}$. In der Bilanz der elektrischen Energie tritt $+\boldsymbol{j} \cdot \boldsymbol{E}$ als Verlustterm auf.

Zur Lösung der Wärmeleitungsgleichung, die eine partielle Differentialgleichung von vier Variablen ist, müssen Anfangsbedingungen $T(\boldsymbol{r}, 0)$ und Randbedingungen vorgegeben werden. Die Randbedingung kann darin bestehen, daß man die Temperatur auf dem Rande vorgibt. Ist das System in ein Wärmebad eingebettet, dann ist die Temperatur auf dem Rand gleich der Temperatur des Wärmebades. Es kann aber auch die Wärmestromdichte bzw. der Temperaturgradient normal zur Oberfläche vorgegeben werden. Für ein wärmeisoliertes System verschwindet die Normalableitung der Temperatur. Die Randbedingungen können schließlich auch in einer Kombination beider Aussagen stehen. Bei der Ableitung der Wärmeleitungsgleichung wurde vorausgesetzt, daß das Material homogen ist. Im anderen Fall ist die Wärmeleitfähigkeit ortsabhängig. In einem stückweise homogenen System ist an den Grenzflächen zu fordern, daß die Normalkomponente der Wärmestromdichte stetig verläuft.

Führt man einem System konstanter Temperatur T_0 kurzzeitig (zur Zeit $t = 0$) in einem kleinen Volumenelement (um $\boldsymbol{r} = 0$) eine Wärmemenge Q zu (Wärmeblitz), dann tritt infolge der Wärmeleitung im Zeitablauf auch eine Erwärmung der Umgebung dieses Ortes auf. Die Wärmeleitungsgleichung liefert die Lösung

$$T(\boldsymbol{r}, t) = \frac{Q}{c\varrho} \frac{1}{(4\pi at)^{3/2}} \mathrm{e}^{-r^2/(4at)}.$$

Die Ausdehnung des erwärmten Gebietes wächst proportional zu $\sqrt{4at}$ an.

Infolge der im Tag-Nacht- oder Sommer-Winter-Zyklus periodischen Erwärmung der Erdoberfläche treten *Temperaturwellen* auf. Wählt man die Randbedingung $T(0, t) = T_0 \cos \omega t$, dann ist $T(x, t) = T_0\, \mathrm{e}^{-x/l} \cos(\omega t - x/l)$ die Lösung für die Schwankungen mit der Amplitude T_0 um die mittlere Temperatur. Für die in der Kreisfrequenz $\omega = 2\pi/\tau$ auftretende Periode τ ist ein Jahr bzw. ein Tag einzusetzen. Die Temperaturwellen klingen nach dem Erdinnern hin ab. Die Eindringtiefe beträgt für die Jahresschwankungen etwa 1 m, für die täglichen Schwankungen nur wenige Zentimeter.

2.6.11.4. Diffusion. Im thermodynamischen Gleichgewicht (bei konstanter Temperatur) ist die Dichte konstant, ortsunabhängig. Diese Aussage ist gleichbedeutend mit der, daß sich ein Druckausgleich einstellt. Natürlich wird dabei nur an ein einphasiges System gedacht, verschiedene Phasen haben im Gleichgewicht verschiedene Dichten. Bei einer mehrkomponentigen Phase sind im thermodynamischen Gleichgewicht aber auch die Dichten ϱ_i der einzelnen Komponenten bzw. die Konzentrationen $x_i = z_i/z$ ortsunabhängig, d. h., auch die Partialdrücke gleichen sich aus. Der Partialdruck gleicht sich langsamer aus als der Gesamtdruck. Dieser Ausgleich erfordert ein Durchdringen der Komponenten, eine Diffusion. Die Diffusionsgeschwindigkeit ist unterschiedlich groß. In Festkörpern kann sie so klein sein, daß sich das thermodynamische Gleichgewicht im Beobachtungszeitraum nicht einstellt. Zur Berechnung der Orts- und Zeitabhängigkeit der Dichteverteilungen beim Einstellen des Gleichgewichtszustandes muß man die Bilanzgleichung $\partial\varrho_i/\partial t + \nabla \boldsymbol{j}_i + \nu_i = 0$ und die Materialgleichung für die Massenstromdichte $\boldsymbol{j}_i = -D \nabla \varrho_i$ heranziehen. Die Massenverlustdichte ν_i ist z. B. bei chemischen Reaktionen von Null verschieden. D wird als *Diffusionskoeffizient* bezeichnet. Sein Wert hängt nicht nur von der diffundierenden Komponente, sondern auch von der Matrix ab. Der Diffusionskoeffizient ist stark temperaturabhängig. In Festkörpern ist er über $D = D_0 \exp[-Q/(RT)]$ mit einer Aktivierungsenergie Q verknüpft (Tab. 2.6.11–5).

Tab. 2.6.11-5 Diffusionskoeffizienten

Substanz		D in cm^2/s
H_2	Selbstdiffusion bei 0 °C und Normaldruck	1,459
O_2		0,191
N_2		0,154

Substanz	D_0 in cm^2/s	Q in kJ/mol
Cu in Ag	1,23	193
Cu in Cu	0,47	197,5
Au in Cu	0,17	194
H in α-Fe[1]	0,12	32,7
H in Pd[2]	0,015	28,5

[1] 25–200 °C und [2] 100–400 °C nach LANDOLT-BÖRNSTEIN

Aus Bilanz- und Materialgleichung folgt die Diffusionsgleichung. Für $\nu_i = 0$ lautet sie

$$\partial\varrho_i/\partial t = D\Delta\varrho.$$

Sie hat eine zur Wärmeleitungsgleichung analoge Form. Daher treten auch analoge Lösungen auf. Eine anfangs punktförmige Anhäufung der Masse M einer Komponente

Tab. 2.6.11–6 Absolute differentielle Thermospannungen Σ bei 0 °C *(oben)* und Thermospannung einiger Thermoelemente bezogen auf 0 °C in mV *(unten)* (nach LANDOLT-BÖRNSTEIN)

Element	Σ in 10^{-6} V/K	Element	Σ in 10^{-6} V/K
Ag	1,4	Fe	17,0
Cu	1,7	Al	−1,6
Au	1,1	K	−15,6
Bi (‖ c)	−100	Li	11,5
Bi (⊥ c)	−53	Pt	−4,4

	−100 °C	0 °C	20 °C	50 °C	100 °C	200 °C	500 °C
Pt/Rh*-Pt	–	0	0,113	0,299	0,643	1,436	4,221
Cu-Konstantan	−3,40	0	0,80	2,05	4,25	9,20	27,40
Fe-Konstantan	−4,60	0	1,05	2,65	5,37	10,95	27,84

	800 °C	900 °C	1 000 °C	1 100 °C	1 200 °C	1 300 °C	1 400 °C
Pt/Rh-Pt	7,329	8,432	9,570	10,741	11,935	13,138	14,337

* 10% Rhodium. Dieses Thermoelement wird als Normal für die Darstellung der Temperaturskala zwischen 903,89 und 1 337,58 K benutzt.

breitet sich im Zeitablauf räumlich entsprechend $\varrho(r, t) = M(4\pi Dt)^{-3/2} \exp[-r^2/(4Dt)]$ aus. Die Ausdehnung der Verteilung wächst proportional zur Wurzel aus der Zeit an.

2.6.11.5. Thermoelektrische Effekte. Bei Thermoelementen wird die Kopplung zwischen kalorischen und elektrischen Erscheinungen ausgenutzt. Besteht in einem Leiterstück ein Temperaturgefälle, dann baut sich in ihm ein elektrisches Feld auf. Zwischen seinen Enden besteht eine Spannungsdifferenz ΔU *(Seebeck-Effekt).* Diese ist zur Temperaturdifferenz $\Delta T = T_2 - T_1$ proportional $\Delta U = \Sigma \Delta T$. Der Proportionalitätsfaktor Σ wird als *differentielle Thermospannung* bezeichnet (Tab. 2.6.11–6). Im *Thermoelement* (Abb. 2.6.11–7) werden zwei verschiedene Metalle (A bzw. B) aneinander gekoppelt. Es liefert die elektromotorische Kraft $U_{AB} = (T_2 - T_1)(\Sigma_A - \Sigma_B)$. Indem man viele Thermoelemente hintereinander schaltet, kann man diese Spannung vervielfachen. Thermoelemente eröffnen z. B. eine Möglichkeit, um Sonnenenergie unmittelbar in elektrische Energie umzuwandeln. Thermoelemente können auch zur Temperaturmessung genutzt werden. Ist T_2 bekannt, dann kann durch Messung der EMK auf T_1 geschlossen werden. Auf diese Weise kann man Temperaturen an relativ unzugänglichen Stellen messen. Die Thermoelemente sind auch bei tiefen Temperaturen, wo die Thermometerflüssigkeit gefrieren würde, einsetzbar.

Wegen der Kopplung zwischen Wärmestrom und elektrischem Strom $I_w = \Pi I$, können Thermoelemente als Kühlelemente eingesetzt werden *(Peltier-Effekt).* $\Pi = T\Sigma$ wird als Peltier-Koeffizient bezeichnet (vgl. Tab. 2.6.11–5). Fließt durch ein Thermoelement ein elektrischer Strom I (Abb. 2.6.11–8), dann kann an den Lötstellen eine konstante Temperatur nur dadurch aufrecht erhalten werden, daß man den Wärmestrom $(\Pi_A - \Pi_B) I$ zu- bzw. abführt. Sonst kühlt sich eine Lötstelle ab, eine erwärmt sich.

Abb. 2.6.11–7 Thermoelement

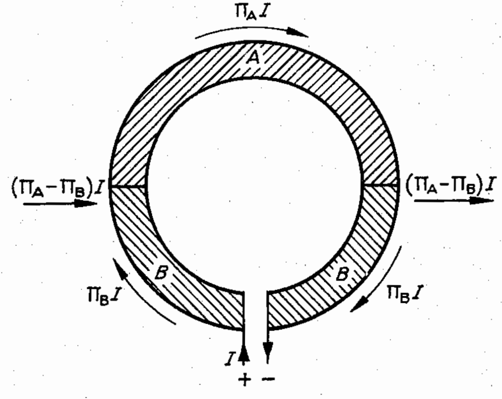


Abb. 2.6.11–8 Peltier-Effekt

In den *Transportgleichungen* zur mathematischen Beschreibung thermoelektrischer Effekte müssen Temperaturgradient und elektrisches Feld als thermodynamische Kräfte betrachtet werden. Sie sind die Ursachen für den elektrischen bzw. Wärmestrom. Da die Reaktionen in dem Material linear von den Kräften abhängen, müssen die Transportgleichungen die Form

$$j = \sigma E - \sigma \Pi \nabla T/T, \quad j_w = \sigma \Pi E - K \nabla T/T = \Pi j - \varkappa \nabla T$$

haben. Zunächst würden vier Koeffizienten auftreten. Der Koeffizient, der den elektrischen Strom infolge eines Temperaturgradienten (hier $\sigma\Pi$ genannt) beschreibt, ist aber nach dem Onsagerschen Reziprokitätstheorem gleich dem Koeffizienten, der den Wärmestrom infolge eines elektrischen Feldes bestimmt. Die thermodynamische Kraft ist in diesem Beispiel $-\nabla T/T$. σ ist die *elektrische Leitfähigkeit.* Sie bestimmt den elektrischen Strom infolge eines elektrischen Feldes in einem Leiter konstanter Temperatur. Die *Wärmeleitfähigkeit* $\varkappa$ wird unter der Bedingung gemessen, daß kein elektrischer Strom fließt ($j = 0$), nicht unter der Bedingung, daß kein elektrisches Feld vorhanden ist. Es gilt $\varkappa = (K - \sigma\Pi^2)/T$, wobei der zweite Summand bei Metallen klein gegen den ersten ist.

Aus der Transportgleichung folgt, daß bei Temperaturunterschieden $\nabla T \neq 0$ auch ohne elektrisches Feld $E = 0$ ein elektrischer Strom fließt. Er ist ∇T entgegen, also zum kälteren Leiterteil hin gerichtet. Ein elektrischer Strom ($j = 0$) fließt nur dann nicht, wenn ein elektrisches Feld der Größe $E = \Pi \nabla T/T = \Sigma \nabla T$ vorhanden ist. Der Temperaturgradient bewirkt einen Dichtegradienten der Leitungselektronendichte. Dadurch wird ein elektrisches Feld aufgebaut. Das ist der *Seebeck-Effekt.* Die elektromotorische Kraft ist $U = \int \mathrm{d}s \cdot E$.

Für den Spezialfall konstanter Temperatur ($\nabla T = 0$) erhält man aus den Transportgleichungen die für den *Peltier-Effekt* wesentliche Beziehung $j_w = \Pi j$. Die *Energiebilanz* führt mit obigen Materialgleichungen auf

$$\partial u/\partial t = \varkappa \Delta T + j^2/\sigma + \tau j \cdot \nabla T.$$

Der erste Summand ist von der Wärmeleitungsgleichung her bekannt. Der zweite Summand stellt die durch den elektrischen Strom erzeugte Joulesche Wärme dar. Der dritte Summand beschreibt die thermoelektrischen Effekte. Der Thomson-Koeffizient $\tau = -T(\mathrm{d}\Sigma/\mathrm{d}T)$ kann beide Vorzeichen annehmen. Er hat für Eisen den Wert $\tau = -15{,}3 \cdot 10^{-6}$ V/K. Seine Wirkung kann am einfachsten beim *Thomson-Effekt* (Abb. 2.6.11–9) beobachtet werden. Ist die Flüssigkeit im Gefäß auf einer höheren Temperatur als der Außenraum, dann zeigen Stromdichte und Temperaturgradient an der Eintauchstelle 1 in die entgegengesetzte, an der Eintauchstelle 2 in gleiche Richtung. Daher kühlt sich für $\tau > 0$ der Draht im Bereich 1 ab, während er bei 2 wärmer wird.

2.6.11.6. Irreversible Prozesse und Strukturbildung. Bei der Diskussion der Transportgleichungen wurde angenommen, daß sich das betrachtete System nahe dem thermodynamischen Gleichgewicht befindet. Die Abweichungen von der Gleichgewichtsverteilung sind klein und beim Anliegen einer äußeren Störung proportional zu dieser Störung. Als Folge dieser Beschränkungen können diese Transportprozesse durch lineare Gleichungen beschrieben werden. Daneben spielen Prozesse weitab vom Thermodynamischen Gleichgewicht eine bedeutende Rolle, insbesondere bei der biologischen Evolution, da bei diesen Prozessen Strukturbildungen möglich sind.

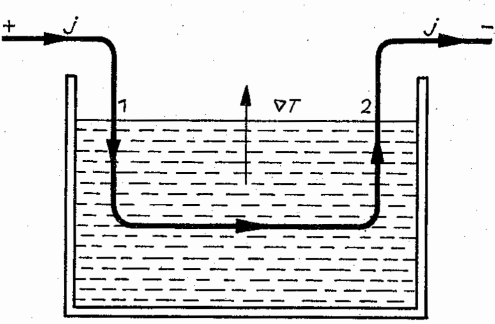


Abb. 2.6.11–9 Thomson-Effekt

In engem Zusammenhang mit dieser Problematik steht der scheinbare Widerspruch zwischen dem zweiten Hauptsatz der Wärmelehre und der zunehmenden Strukturiertheit des Lebens auf der Erde. Nach dem zweiten Hauptsatz kann die Entropie in einem abgeschlossenen System nur anwachsen. Da auf der Erde – wie auch auf anderen Himmelskörpern – ständig irreversible Prozesse ablaufen, wird in hohem Maße Entropie produziert. Davon ausgehend wurde die Behauptung aufgestellt, daß wir unweigerlich einem Chaos, dem Wärmetod, entgegengehen. Dem muß man entgegenhalten, daß der zweite Hauptsatz nicht auf den Kosmos als Ganzes angewendet werden kann. Es ist ein Erfahrungsgesetz, das aus Beobachtungen in einem kleinen Teil des Kosmos gefunden wurde und daher nicht ohne weiteres verallgemeinert werden kann. Wir wissen z. B., daß neue Sternsysteme im Kosmos entstehen. Die damit zusammenhängenden Vorgänge sind aber weitgehend unbekannt.

Was die Erde als Teilsystem betrifft, so findet auf ihr zwar eine erhebliche Entropieproduktion statt, die Erde ist aber

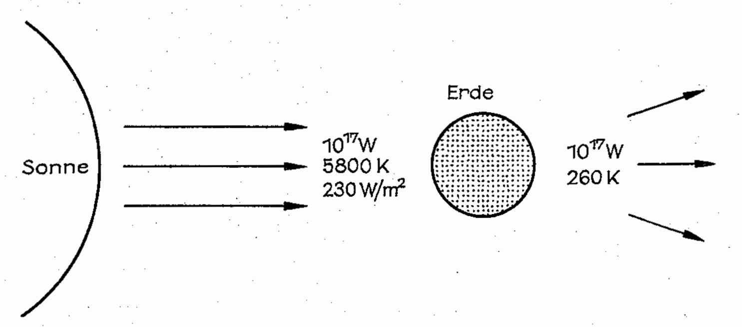


Abb. 2.6.11–10 Entropieexport der Erde (nach Ebeling),

$$\Delta S = \frac{10^{17}\,\mathrm{W}}{5\,800\,\mathrm{K}} - \frac{10^{17}\,\mathrm{W}}{260\,\mathrm{K}}$$

kein abgeschlossenes System. Durch die Aufnahme der Sonnenstrahlung und die eigene Abstrahlung (Abb. 2.6.11–10) realisiert die Erde einen enormen Entropieexport, der die Entropieproduktion auf der Erde übersteigt. Die Energiebilanz ist dagegen etwa ausgeglichen. Dieser Entropieexport – und damit letztlich die Sonnenstrahlung – ermöglicht die zunehmende Strukturiertheit des Lebens auf der Erde.

An einfachen Modellbeispielen gibt es Untersuchungen, um die Bedingungen dafür zu finden, daß sich neue Strukturen durchsetzen. Es hat sich eindeutig herausgestellt, daß diese Strukturen nicht das Ergebnis von Schwankungserscheinungen um den thermodynamischen Gleichgewichtszustand sind. Die Strukturbildung tritt bei irreversiblen Prozessen weitab vom Gleichgewicht auf. In der mathematischen Beschreibung dieser Prozesse treten Lösungen von nichtlinearen Gleichungen auf, die bei bestimmten Parameterwerten ineinander umschlagen. Bildet sich eine neue Struktur, so ist sie als Störung des vorhandenen Zustandes anzusehen. Ist das System stabil gegen die neue Struktur, so verschwindet sie wieder, im anderen Falle stabilisiert sie sich. Der sich einstellende stationäre Zustand ist dadurch gekennzeichnet, daß die Entropieproduktion minimal ist.

Ein bekanntes Beispiel für eine Strukturbildung ist der *Bernard-Effekt* (Abb. 2.6.11–11). Durch Erwärmen des unteren Bereiches einer Ölschicht werden zunächst Diffussionsprozesse in Gang gesetzt. Die Wärmeleitung versucht den Zustand, in dem sich die Schichten mit geringerem spezifischem Gewicht unten befinden, abzubauen. Bei wachsender Temperaturdifferenz wird dieser Prozeß instabil. Es tritt ein Punkt auf, wo er in einen Prozeß mit Wärmekonvektion umschlägt. Es bilden sich – weitgehend unabhängig von der Form der Berandung der Ölschicht – sechseckige Bereiche heraus, in denen die Flüssigkeit im Inneren aufsteigt und im Außenbereich absinkt.

Ein Umschlagen in eine andere Struktur setzt Strukturreserven, die Möglichkeit, andere Strukturen anzunehmen, voraus. Biologische Objekte haben große Strukturreserven, durch Mutationen werden laufend neue Strukturen gebildet. Je nach dem, ob diese Strukturen der Umwelt besser angepaßt sind oder nicht, setzen sie sich durch, oder sie werden wieder zerstört. Auch der Mensch realisiert einen beträchtlichen Entropieexport. Die Nahrung mit ihren Molekülen hoher Struktur hat eine relativ kleine Entropie. Die Ausscheidungen mit ihren einfachen Molekülen haben eine höhere Entropie. Dadurch wird die Lebensfähigkeit, die Bildung von hochstrukturierten Systemen wie auch der Ablauf der irreversiblen Transportprozesse im Körper, gewährleistet. Die Energiebilanz ist dagegen im wesentlichen ausgeglichen.

2.6.12. Kinetische Gastheorie

Die Wärmelehre mit ihren Hauptsätzen als Kernstück ist eine phänomenologische Theorie. Die individuelle Bewegung des einzelnen Atoms oder Moleküls wird nicht untersucht. Das System wird durch wenige Parameter, wie Druck, Temperatur und Volumen, beschrieben. Ein spezielles System wird durch eine spezielle Zustandsgleichung bzw. ein spezielles thermodynamisches Potential charakterisiert. In der statistischen Physik wird der Zusammenhang zwischen der individuellen Bewegung der Bausteine des Systems und seinem makroskopischen Verhalten hergestellt. Sie ermöglicht es, Materialkonstanten, wie Kompressibilität oder spezifische Wärme, von den zwischenatomaren Kräften ausgehend zu berechnen und Zusammenhänge zwischen diesen Koeffizienten aufzudecken, die aus deren mikroskopischer Begründung folgen.

2.6.12.1. Boltzmann-Wahrscheinlichkeit. Betrachtet man ein System, das mit einem Wärmebad der Temperatur T in Kontakt steht, so hat es im thermodynamischen Gleichgewicht auch die Temperatur T angenommen. Dies ist kein statischer Zustand. Die Atome stoßen auf die Gefäßwände, wobei sie in ungeordneter Weise Energie verlieren oder aufnehmen. Die mechanische Energie ist also gewissen Schwankungen unterworfen. Dies wird in der statistischen Physik durch eine Wahrscheinlichkeitsaussage ausgedrückt.

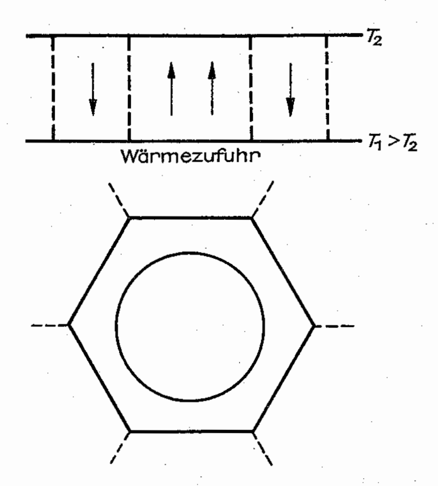


Abb. 2.6.11–11 Bernard-Effekt, unten Draufsicht

Die grundlegende Aussage der statistischen Physik, die die Verbindung zur phänomenologischen Beschreibung herstellt, besteht in dem Folgenden: Ein System, das mit einem Wärmebad der Temperatur T in Kontakt steht, befindet sich mit der Wahrscheinlichkeit *(Boltzmann-Wahrscheinlichkeit)*

$$W_B = C\,e^{-E/kT}$$

in einem Zustand mit der Energie E. k nennt man die *Boltzmann-Konstante*, $k = 1{,}381 \cdot 10^{-23}$ J/K. C ist ein Normierungsfaktor, der so bestimmt werden muß, daß die Wahrscheinlichkeit, das System mit irgendeinem Wert der Energie anzutreffen Eins ist. Je kleiner die Energie, desto größer ist die Boltzmann-Wahrscheinlichkeit. Man darf daraus nicht schließen, daß sich ein System die meiste Zeit in den Zuständen kleinster Energie aufhält. Schon bei der Diskussion der Maxwellschen Geschwindigkeitsverteilung wird es deutlich werden, daß die Zahl der Zustände, die es zu einer Energie gibt, mit der Energie stark anwächst. Im Zusammenspiel beider Tendenzen wird eine, von der Temperatur abhängige, wahrscheinlichste Energie ausgezeichnet.

2.6.12.2. Maxwellsche Geschwindigkeitsverteilung. Die Atome eines (fast) idealen Gases ändern durch die Stöße untereinander und mit den Gefäßwänden ständig ihre Energie, die in diesem Fall gleich der kinetischen Energie $(m/2)\,v^2$ ist. Die Wahrscheinlichkeit, im thermodynami-

schen Gleichgewicht ein Atom mit einer Geschwindigkeit $\boldsymbol{v}$ im (Geschwindigkeits-)Volumenelement $d^3\boldsymbol{v} = dv_x dv_x dv_z$ anzutreffen, ist durch die Boltzmann-Wahrscheinlichkeit

$$W(\boldsymbol{v})\,d^3\boldsymbol{v} = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{mv^2/2}{kt}}\,d^3\boldsymbol{v}, \quad \int W(\boldsymbol{v})\,d^3\boldsymbol{v} = 1$$

gegeben. Sie hängt nur vom Betrag der Geschwindigkeit ab, und wird mit wachsender Geschwindigkeit kleiner. Markiert man im Geschwindigkeitsraum jedes Atom durch den Wert seiner Geschwindigkeit, dann erhält man das in Abb. 2.6.12–1 skizzierte Bild. Fragt man nach der Wahrscheinlichkeit $W_1(v)\,dv$, ein Atom mit einem bestimmten

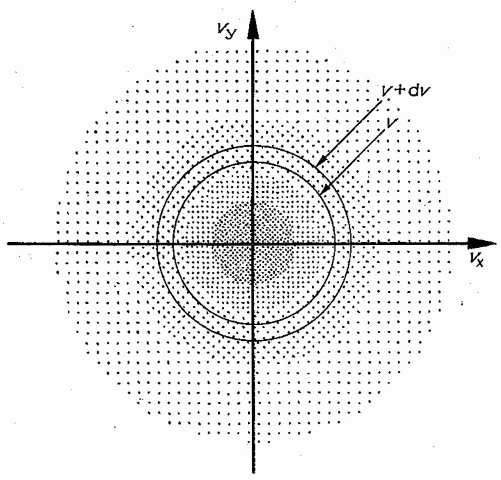


Abb. 2.6.12–1 Geschwindigkeitsverteilung im Geschwindigkeitsraum dargestellt

Wert des Betrages der Geschwindigkeit anzutreffen, dann muß man alle Atome innerhalb einer Kugelschale (vgl. Abb. 2.6.12–1) mit den Radien v und $v + dv$ und dem Volumen $4\pi v^2 dv$ in Betracht ziehen,

$$W_1(v)\,dv = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-\frac{mv^2/2}{kT}}\,4\pi v^2 dv, \quad \int W_1(v)\,dv = 1\,.$$

Diese Verteilung (Abb. 2.6.12–2) hat ein Maximum bei der „wahrscheinlichsten" Geschwindigkeit $v_w = \sqrt{2kT/m}$.
Von den Wahrscheinlichkeiten $W(\boldsymbol{v})$ oder $W_1(v)$ ausgehend können *Geschwindigkeitsmittelwerte* berechnet werden:
$\sqrt{\overline{v^2}} = \sqrt{3kT/m} = \left[\int dv\,v^2\,W_1(v)\right]^{1/2}$,
$\bar{v} = \sqrt{(8/\pi)\,kT/m} = \int dv\,v W_1(v) = \sqrt{\overline{v^2}}/1{,}086$.
Von Interesse ist auch der Betrag der *mittleren Relativgeschwindigkeit* zweier Atome $\bar{v}_r = |\overline{\boldsymbol{v}_1 - \boldsymbol{v}_2}| = \bar{v}/\sqrt{2}$. Die Größenordnung der Atom- oder Molekülgeschwindigkeiten ergibt sich aus, $\bar{v} = 145\,\sqrt{(T/\mathrm{K})/M}$ m/s. Bei Zimmertemperatur $T = 293$ K erhält man für ein Wasserstoffmolekül (relative Molekülmasse $M = 2$) $\bar{v} = 1\,755$ m/s, für ein Sauerstoffmolekül ($M = 32$) $\bar{v} = 439$ m/s.

2.6.12.3. Gleichverteilungssatz. Aus dem Mittelwert $\overline{v^2} = 3kT/m$ kann unmittelbar auf die kinetische Energie eines Atoms geschlossen werden,

$$\overline{W}_{kin} = (3/2)\,kT.$$

Dieses Ergebnis ist eine spezielle Form der allgemeinen Aussage der klassischen Statistik: Die *mittlere kinetische Energie* je Freiheitsgrad beträgt im thermodynamischen Gleichgewicht $kT/2$ (Gleichverteilungssatz). Daraus ist unmittelbar ersichtlich, daß die innere Energie eine Funktion der Temperatur ist und daß die Temperatur ein Maß für die ungeordnete Wärmebewegung der Atome und Moleküle darstellt.
Die Wärmebewegung der Atome beeinflußt andere physikalische Größen. Um einen Eindruck von der Stärke dieses Einflusses zu bekommen, muß die Energie kT mit anderen charakteristischen Energien verglichen werden. So treten bei Elektronenanregungen Energien von der Größe eV (Elektronenvolt, die Energie, die einem Elektron beim Durchlaufen des Spannungsgefälles 1 V übertragen wird) auf. Damit kT die Größe 1 eV hat, muß die Temperatur etwa $T = 10^4$ K betragen. Bei Zimmertemperatur hat kT also die Größe 0,03 eV. In welchem Maße in einem Halbleiter Elektronen durch die Wärmebewegung vom Valenzband in das Leitungsband angeregt werden, hängt vom Verhältnis dieser Energie zur Breite der Energielücke ab.
Aus dem Gleichverteilungssatz folgt, daß die innere Energie U eines *einatomigen idealen Gases* je Mol $U = L(3/2)\,kT = (3R/2)\,T$ ist, die Molwärme also $C_V = 3R/2$. Bei wechselwirkenden Teilchen muß die *potentielle Energie* in der inneren Energie berücksichtigt werden. Der *Virialsatz* verknüpft die zeitlichen Mittelwerte von potentieller und kinetischer Energie gebundener Zustände. Aus ihm folgt für eine harmonische Kraft

$$\overline{W}_{pot} = \overline{W}_{kin} \quad \text{und damit} \quad U = \overline{W}_{kin} + \overline{W}_{pot} = 2\,\overline{W}_{kin}.$$

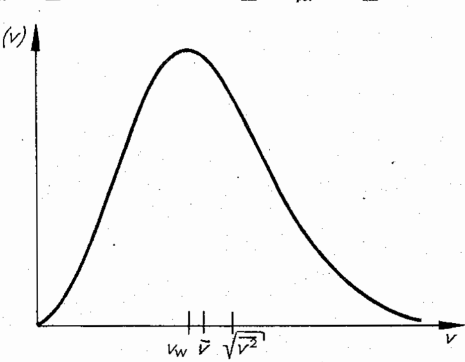


Abb. 2.6.12–2 Maxwellsche Geschwindigkeitsverteilung $W_1(v)$

Ein *zweiatomiges Molekül* mit starrer Verbindung zwischen den Atomen hat fünf Freiheitsgrade, also die Molwärme $C_V = 5R/2$. Ist die Relativbewegung – mit der bei kleinen Auslenkungen aus der Gleichgewichtslage harmonischen Kraft – am thermodynamischen Gleichgewicht beteiligt, dann sind kinetische und potentielle Energie dieses sechsten Freiheitsgrades zur inneren Energie hinzuzufügen. Die Molwärme wird also $C_V = 7R/2$. Man beobachtet bei einem Gas zweiatomiger Moleküle mit ansteigender Temperatur einen Übergang von $C_V = 3R/2$ über $C_V = 5R/2$ nach $C_V = 7R/2$. Dies ist darin begründet, daß bei tiefen Temperaturen die klassische Statistik – und damit auch der Gleichverteilungssatz – nicht mehr gültig ist. Die Quantenstatistik beschreibt diese Temperaturabhängigkeit richtig.
Für einen *Festkörper*, in dem die Atome bei kleinen Auslen-

kungen harmonische rücktreibende Kräfte spüren, folgt aus dem Virialsatz $\overline{W}_{pot} = \overline{W}_{kin}$ und aus dem Gleichverteilungssatz $\overline{W}_{kin} = (3R/2)\,T$ (je Mol). Dies erklärt die *Dulong-Petitsche Regel* $C_V = 3R$, die für hohe Temperaturen (oberhalb der Debye-Temperatur) beobachtet wird.

2.6.12.4. Gasdruck. Von den Vorstellungen der kinetischen Gastheorie ausgehend kann man leicht den Druck eines idealen Gases auf die Gefäßwand und damit die Zustandsgleichung bestimmen. Bei einer Reflexion an einer glatten Wand (Abb. 2.6.12–3) wird die x-Komponente des Impulses im Vorzeichen umgekehrt, also der Impuls $2p_x$ auf die Wand übertragen. In der Zeit Δt übertragen alle Atome (der betrachteten Geschwindigkeit) aus einem Zylinder der Höhe $v_x \Delta t$ einen solchen Impuls auf die Fläche ΔA. Da die Impulsänderung ΔP gleich dem Kraftstoß $F\Delta t$ ist, ergibt sich der Druck aus $\Delta P/\Delta A\,\Delta t$. Man muß noch die Beiträge der Atome verschiedener Geschwindigkeit addieren, um die gesamte Impulsübertragung zu erhalten. Da p_x und Zylindervolumen beide proportional zu v_x sind, entspricht dies einer Mittelwertbildung über v_x^2, die zu einem zu T proportionalen Wert führt. Man erhält $pV = zRT$, die Zustandsgleichung des idealen Gases.

2.6.12.5. Barometrische Höhenformel. Ein Teilchen im Erdschwerefeld hat die Energie $E = (m/2)\,v^2 + mg\,(h - h_0)$, wenn die potentielle Energie in der Höhe h_0 gleich Null gesetzt wird. Die Boltzmann-Wahrscheinlichkeit $W_B \sim e^{-E/kT}$ beinhaltet hier die Aussage, daß mit wachsender Höhe die Wahrscheinlichkeit abnimmt, ein Teilchen anzutreffen. Die Dichte ist also entsprechend $\varrho = \varrho_0\, e^{-mg(h-h_0)/kT}$ höhenabhängig. Dabei ist ϱ_0 die Dichte in der Höhe h_0. Einen solchen Dichteverlauf beobachtet man z. B. bei Aufschwemmungen. In einem Gemisch von Teilchen verschiedener Masse nimmt die Dichte der schweren Teilchen rascher mit der Höhe ab. Dies kann zur *Trennung der Teilchensorten* genutzt werden. Die Trennung ist um so vollständiger möglich, je größer der Unterschied mg/kT für die Bestandteile ist. In der *Zentrifuge* wird statt der Schwerebeschleunigung die Zentrifugalbeschleunigung genutzt. In einer Ultrazentrifuge können Beschleunigungen bis zum 10^6-fachen Wert der Erdbeschleunigung erzielt werden. Isotope, die sich nur in ihrer Masse, nicht in ihren chemischen Eigenschaften unterscheiden, lassen sich in ihr trennen.

Die barometrische Höhenformel wurde zur Bestimmung der *Loschmidtschen Konstanten* genutzt. Bei bekannter Masse m läßt sich nach Auszählen der Teilchenzahlen in verschiedener Höhe die Boltzmann-Konstante k bzw. $L = R/k$ berechnen. Dies wurde erstmals von PERRIN durchgeführt, der die Dichte von Schwebeteilchen in Wasser in verschiedenen Höhen bestimmte. Die aus Dichte und Größe bekannte Masse der Teilchen betrug etwa 10^{-16} kg.

2.6.12.6. Mittlere freie Weglänge. Zwischen den Atomen oder Molekülen wirken Kräfte, so daß die Moleküle eines Gases Stöße untereinander ausführen. Einem Molekül steht im Mittel das Volumen $V/N = 1/n \sim d^3$ zur Verfügung (n = Teilchendichte). Ist die Reichweite der zwischenmolekularen Kräfte klein gegen d, dann kann man das Bild entwerfen, daß ein Molekül eine gewisse Strecke mit konstanter Geschwindigkeit fliegt, mit einem Molekül zusammenstößt und seine Geschwindigkeit ändert, wieder eine Strecke mit konstanter Geschwindigkeit fliegt usw.

Ein Molekül stellt dabei eine Hindernisfläche dar, die durch seinen Wirkungsquerschnitt σ (vgl. 2.1.10.4.) gegeben ist. Die mittlere Relativgeschwindigkeit zweier Moleküle $\bar{v}_r = \bar{v}/\sqrt{2}$ ist bekannt. Die Hindernisfläche überstreicht während einer Zeit τ das Volumen $\sigma \bar{v}_r \tau$. Im Mittel findet ein Stoß statt, wenn dieses Volumen gleich dem Volumen je Teilchen $1/n$ ist. Die *mittlere Stoßzeit* ist also durch $\tau = 1/(\sigma n \bar{v}_r)$ gegeben. Als *mittlere freie Weglänge* l bezeichnet man die Strecke, die ein Teilchen in dieser Zeit zurücklegt,

$$l = \tau\bar{v} = 1/(\sqrt{2}\,\sigma n).$$

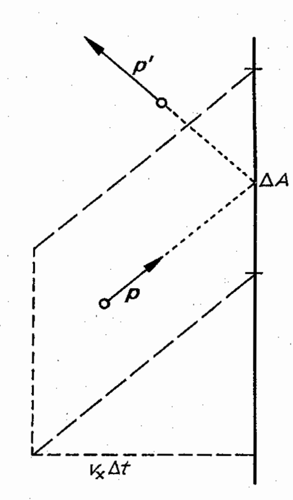


Abb. 2.6.12–3 Druck durch Stöße auf eine Fläche ΔA

Der Wirkungsquerschnitt und damit die mittlere freie Weglänge sind energieabhängig. Da die mittlere Energie der stoßenden Partner im thermodynamischen Gleichgewicht durch die Temperatur bestimmt ist, müssen diese Größen als temperaturabhängig angesehen werden. In Gasen ist unter Normalbedingungen l von der Größenordnung 10^{-5} cm. Je Sekunde werden etwa 10^{10} Stöße ausgeführt, d. h. $\tau \approx 10^{-10}$ s. Im Vergleich dazu ist bei Normalbedingungen die lineare Ausdehnung des mittleren Volumens je Molekül $d = 0{,}334 \cdot 10^{-8}$ m.

2.6.12.7. Transportkoeffizienten in Gasen. Mittels der Begriffe mittlere Geschwindigkeit und mittlere freie Weglänge kann man die in 2.6.11. eingeführten Transportkoeffizienten auf mikroskopische Prozesse zurückführen und einfache Näherungsformeln gewinnen.

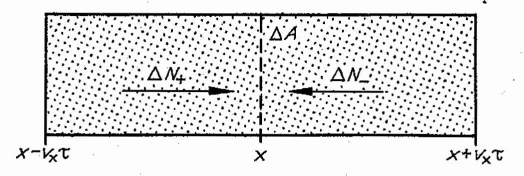


Abb. 2.6.12–4 Teilchenströme durch ΔA

Diffusionskoeffizient. Er verbindet die Teilchenstromdichte mit dem Gefälle der Teilchendichte $j = -D\,dn/dx$ (wenn x die Richtung stärksten Gefälles, d. h. des Gradienten, ist). Die durch eine Fläche ΔA während der mittleren Stoßzeit τ hindurchtretenden Teilchenströme (Abb. 2.6.12–4) sind durch das Zylindervolumen $\Delta A\, v_x \tau$ und die mittlere Dichte in den Zylindern $n(x - v_x\tau/2)$ bzw. $n(x + v_x\tau/2)$ bestimmt, $j\Delta A\tau = \Delta N_+ - \Delta N_- = \Delta A\, v_x \tau[n(x - v_x\tau/2) - n(x + v_x\tau/2)] = \Delta A\,(v_x\tau)^2\, dn/dx$. Führt man

nach der Mittelwertbildung über alle Teilchengeschwindigkeiten die mittlere freie Weglänge ein, dann erhält man

$$D = \bar{v}\, l/3 .$$

Wichtig ist das Ergebnis, daß der Diffusionskoeffizient proportional zur mittleren Geschwindigkeit und zur mittleren freien Weglänge ist und daher auch deren Temperatur- und Druckabhängigkeit übernimmt. Der Faktor 1/3 ist eine Abschätzung.

Wärmeleitfähigkeit. Sie kann aus analogen Überlegungen gewonnen werden. Beim Wärmestrom ist die von den Teilchen transportierte Energie aufzusummieren. Bei einem Temperaturgefälle ist diese für die Teilchenströme ΔN_+ und ΔN_- (vgl. Abb. 2.6.12–4) verschieden groß, weil sich die mittlere Energie je Teilchen unterscheidet. Es ergibt sich

$$\varkappa = \varrho c_V \bar{v} l/3 = \varrho c_V D .$$

Das Produkt von Massendichte ϱ und spezifischer Wärme c_V ist gleich der Wärmekapazität je Volumen oder gleich dem Quotienten von Molwärme zu Molvolumen C_V/V_m. Da der Diffusionskoeffizient zur Dichte umgekehrt proportional ist, wird die Wärmeleitfähigkeit unabhängig von der Dichte. Bei konstanter Temperatur ist sie daher unabhängig vom Druck. Dieser Schluß gilt nicht mehr, wenn im Hochvakuum die freie Weglänge größer als die Gefäßabmessungen wird. Dann steigt die Wärmeleitung mit der Dichte an. Dies ist der Arbeitsbereich des *Wärmeleitungsmanometers.*

Zähigkeit. Sie beschreibt in einem strömenden Gas die Scherspannung oder Impulsstromdichte zwischen Schichten unterschiedlicher Geschwindigkeit $j_{xz} = -\eta\, \mathrm{d}v_x/\mathrm{d}z$ (Abb. 2.6.12–5). Sie beschreibt einen Diffusionsprozeß senkrecht zur Strömungsrichtung. Man erhält für die Zähigkeit

$$\eta = (1/3)\, \varrho \bar{v} l = \varrho D .$$

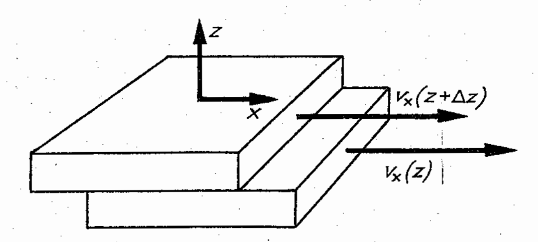


Abb. 2.6.12–5 Scherspannung durch die unterschiedliche Geschwindigkeit benachbarter Schichten

Sie ist wie die Wärmeleitfähigkeit unabhängig vom Gasdruck.

Die aus diesen einfachen gaskinetischen Überlegungen gewonnenen Beziehungen zwischen Transportkoeffizienten, Molwärme und Dichte werden in ihrer Proportionalität experimentell weitgehend bestätigt. In den Zahlenfaktoren sind Korrekturen anzubringen.

2.6.12.8. Schwankungserscheinungen. Für das thermodynamische Gleichgewicht wurden Mittelwerte berechnet, wie Druck, mittlere kinetische Energie usw. Aus den gaskinetischen Betrachtungen wird klar, daß Fluktuationen um diese Mittelwerte auftreten. So wird z. B. die Kraft auf ein Flächenelement um so stärkeren Schwankungen unterworfen sein, je kleiner die Fläche ist, je weniger Atome in einer gegebenen Zeitspanne an der Fläche reflektiert werden.

Bei der *Brownschen Bewegung* führen solche Schwankungen zu makroskopisch sichtbaren Effekten. Beobachtet man ein kleines Teilchen, das sich in einer ruhenden Flüssigkeit befindet, unter dem Mikroskop, dann stellt man eine Zitterbewegung fest, die über längere Zeiten zu einem unregelmäßigen Wandern des Teilchens führt. Dies ist so zu verstehen, daß das Teilchen nicht unter einem konstanten hydrostatischen Druck steht, wegen der Kleinheit seiner Oberfläche sind Schwankungen bei der Kraftübertragung durch die Stöße mit den Flüssigkeitsmolekülen registrierbar. Das Teilchen nimmt am thermodynamischen Gleichgewicht teil und hat eine mittlere kinetische Energie $3kT/2$ bezüglich seiner Schwerpunktsbewegung. Für das Quadrat der Strecke Δx, die im Mittel in der Zeit t zurückgelegt wird, ergibt sich $(\Delta x)^2 = 2Dt$. Die Schwankungserscheinungen begrenzen auch die *Empfindlichkeit von Instrumenten.* In einem Spiegelgalvanometer z. B. führt der Spiegel infolge der thermischen Bewegung kleine unregelmäßige Schwingungen aus. Er hat eine mittlere kinetische Energie $kT/2$. Die Auslenkung des Spiegels durch elektrische Ströme, die mit einem Energieumsatz von der Größenordnung kT verbunden sind, kann daher nicht von den thermischen Schwankungen unterschieden werden. Solch kleine Ströme sind mit dem Spiegelgalvanometer nicht mehr meßbar. Die zugehörige Stromstärke liegt bei etwa 10^{-12} A. Das Widerstandsrauschen ist ein anderes Beispiel für den Einfluß von Schwankungserscheinungen.

2.6.13. Dritter Hauptsatz der Thermodynamik

Der dritte Hauptsatz der Thermodynamik, auch *Nernstscher Wärmesatz* genannt, liefert Aussagen über das Verhalten thermodynamischer Größen im thermodynamischen Gleichgewicht bei Annäherung an den absoluten Nullpunkt. Er läßt sich in der Aussage

$$\lim_{T \to 0} S(T, V) = 0 , \quad \lim_{T \to 0} S(T, p) = 0$$

zusammenfassen, daß die Entropie bei Annäherung an den absoluten Nullpunkt einem konstanten, von Volumen, Druck o. a. Zustandsgrößen unabhängigen Wert S_0 zustrebt. Die Quantenstatistik liefert $S_0 = 0$.

Man beachte, daß diese Aussage nur für das thermodynamische Gleichgewicht gilt. Der Gleichgewichtszustand stellt sich bei tiefen Temperaturen aber oft nicht ein, weil erforderliche Platzwechselvorgänge zu langsam vonstatten gehen. Es wird ein metastabiler Zustand eingefroren. Dieser hat eine von Null verschiedene Entropie.

Eine andere Form des dritten Hauptsatzes ist die Aussage von der *Unerreichbarkeit des absoluten Nullpunktes.* Abb. 2.6.13–1 zeigt die Funktion $S(V, T)$ eines angenommenen Systems für zwei Volumina $V_2 > V_1$. Man versucht, das System durch eine Folge von isothermen Kompressionen und adiabatischen Expansionen abzukühlen, wobei sein Volumen zwischen den Werten V_2 und V_1 hin- und herpendeln soll. Da nach dem dritten Hauptsatz die Kurven $S(V_2, T)$ und $S(V_1, T)$ für $T \to 0$ zusammentreffen, ist der absolute Nullpunkt in einer endlichen Anzahl von Schritten nicht erreichbar, also unerreichbar. Für $T \to 0$ berühren sich innere Energie und freie Energie wegen $(U - F)/T = S \to 0$ in höherer Ordnung. Spannungskoeffizient $(\partial p/\partial T)_V \to 0$ und Ausdehnungskoeffizient $(\partial V/\partial T)_p \to 0$

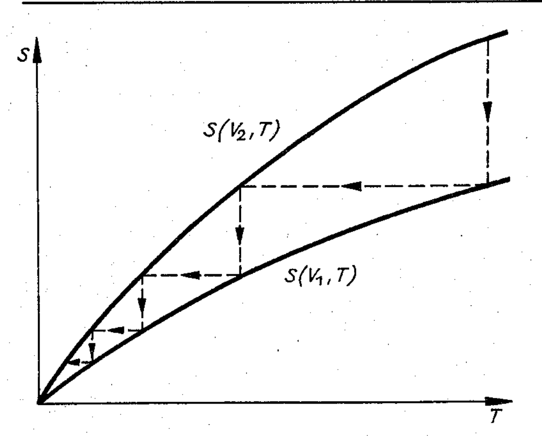


Abb. 2.6.13–1 Abkühlung durch eine Folge von isothermen und adiabatischen Prozessen

verschwinden, ebenso die Molwärmen $C_V \to 0$ bzw. $C_p \to 0$. Wegen $\Delta Q = T\Delta S$ verschwinden mit ΔS auch Umwandlungswärmen. Dies findet seine experimentelle Bestätigung z. B. im Zustandsdiagramm von Helium (vgl. Abb. 2.6.7–3), in dem sich die Schmelzdruckkurve für $T \to 0$ einer Waagerechten nähert. Die Änderung des Schmelzdruckes mit der Temperatur wird aber in Analogie zur Clausius-Clapeyronschen Dampfdruckformel durch die Umwandlungswärme bestimmt.

Die Absolutwerte der Entropie gehen in die Konstante des Massenwirkungsgesetzes ein. Da die Entropie am absoluten Nullpunkt verschwindet, können die Funktionen $S(T, V) = \int_0^T \mathrm{d}T(C_V/T)$ bzw. $S(T, p) = \int_0^T \mathrm{d}T(C_p/T)$ durch Messung der Molwärme $C_V(T)$ bzw. $C_p(T)$ bestimmt werden, ohne daß eine Konstante offenbleibt. Die Zustandsgrößen des idealen Gases erfüllen die Forderungen des dritten Hauptsatzes nicht. Das ideale Gas ist für $T \to 0$ ein ungeeignetes Modell. Man spricht von einer auftretenden Gasentartung, die in der Quantenphysik erfaßt wird.

3. Relativitätstheorie und Gravitation

3.1. Spezielle Relativitätstheorie

3.1.1. Theoretische und experimentelle Situation der Physik vor Aufstellung der Speziellen Relativitätstheorie

Das theoretische Denken war bis 1905 beherrscht durch das Begriffssystem der Newtonschen Mechanik (vgl. 2.1.). Dieses Begriffssystem setzt die Existenz eines dreidimensionalen euklidischen Raumes, einer für alle Beobachter verbindlichen (absoluten) Zeit und einer unendlich großen Ausbreitungsgeschwindigkeit von Wirkungen voraus.

3.1.1.1. Galilei-Transformationen und Relativitätsprinzip der klassischen Mechanik. Die Transformationen, die die Gesetze der Newtonschen Mechanik ungeändert lassen, sind die Galilei-Transformationen $\boldsymbol{r}' = \boldsymbol{r} + \boldsymbol{v}_0 t$, $t' = t$ (absolute Zeit), $\boldsymbol{v}_0$ = konst. Sie führen vom Inertialsystem K in das neue Inertialsystem K'. Eine Folge der Galilei-Transformationen ist das Newtonsche Additionstheorem der Geschwindigkeiten $\boldsymbol{v}' = \boldsymbol{v} + \boldsymbol{v}_0$, wobei $\boldsymbol{v}$, $\boldsymbol{v}'$ Geschwindigkeiten des Massenpunktes in K bzw. K' und $\boldsymbol{v}_0$ die Relativgeschwindigkeit von K und K' sind.

Das Relativitätsprinzip der klassischen Mechanik lautet: Es ist nicht möglich, durch rein mechanische Versuche ein besonderes Inertialsystem auszuzeichnen. Es sind nur Relativgeschwindigkeiten meßbar. Eine Absolutgeschwindigkeit (etwa gegen den Raum oder einen „Äther") ist im Rahmen der Mechanik weder meßbar noch sinnvoll definierbar. Wenn die Beschränkung dieses Prinzips auf die Mechanik wegfällt, kommt man zur Speziellen Relativitätstheorie.

3.1.1.2. Elektrodynamik und ausgezeichnetes Bezugssystem. Vor 1905 glaubte man, daß die Maxwell-Gleichungen in der üblichen Form (vgl. 3.6.2.1.) nur in einem ausgezeichneten Bezugssystem gelten. Dieses System sollte durch den sogenannten Äther, das hypothetische Ausbreitungsmedium des Lichtes, bestimmt sein. Da in diesem „Äthersystem" die Lichtausbreitung im Vakuum richtungsunabhängig ist, wird im bewegten System nach dem Newtonschen Additionstheorem $\boldsymbol{c}' = \boldsymbol{c} - \boldsymbol{v}$ richtungsabhängig. Durch Messung dieser Richtungsabhängigkeit glaubte man, Betrag und Orientierung der „absoluten Geschwindigkeit" (gegenüber dem Äther) für jeden Beobachter angeben zu können und damit die Lage des „Äthersystems" zu bestimmen.

3.1.1.3. Experimente zum Nachweis eines ausgezeichneten Bezugssystems. Durch das Zusammenspiel von Galilei-Transformationen und Elektrodynamik sollte ein ausgezeichnetes Bezugssystem auffindbar werden. Diese Hoffnung erfüllte sich nicht.

Induktion. In einem Leitersystem hängt die induzierte Spannung nur von der Relativgeschwindigkeit gegenüber dem Magnetfeld ab.

Doppler-Effekt. In der klassischen Physik ergeben sich unterschiedliche Frequenzverschiebungen, je nachdem ob sich Lichtquelle oder Beobachter gegenüber dem Äther bewegen. $\lambda' = \lambda(1 - v\cos\vartheta/c)$ bewegte Quelle, $\lambda' = \lambda/(1 - v\cos\vartheta/c)$ bewegter Beobachter, ϑ ist dabei der Winkel zwischen Geschwindigkeit und Ausbreitungsrichtung. Die Relativitätstheorie liefert in beiden Fällen das gleiche Resultat, $\lambda' = \lambda(1 - v\cos\vartheta/c)/\sqrt{1 - v^2/c^2} \approx \lambda(1 - v/c\cos\vartheta + (1/2)(v/c)^2 + \ldots)$. Durch Messungen an Kanalstrahlen konnte von Ives, Stilwell und Otting der quadratische Term in v/c, den die klassische Physik nicht enthält, nachgewiesen werden.

Michelson-Experiment. Ein Lichtstrahl (Abb. 3.1.1–1) fällt auf einen halbdurchlässigen Spiegel. Seine beiden Teile durchlaufen die Wege I und II, werden reflektiert und im Fernrohr zur Interferenz gebracht. Die Apparatur kann als Ganzes gedreht werden, wobei einmal der Weg I, einmal der Weg II in Richtung der Erdbewegung fällt. Nach dem Newtonschen Additionstheorem müßte sich bei dieser Rotation das Streifensystem um $\delta \approx [(l_1 + l_2)/\lambda](v^2/c^2)$ verschieben, so daß man direkt Glieder in $(v/c)^2$ mit großer Genauigkeit messen sollte. Es wurde nicht die geringste Streifenverschiebung beobachtet.

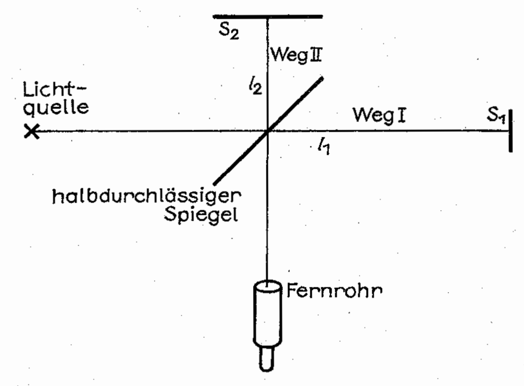


Abb. 3.1.1–1 Michelson-Experiment

Versuch von Trouton und Noble. Ein aufgeladener Kondensator bewegt sich mit der Erde durch den Äther. Seine Platten sind gegen die Bewegungsrichtung schräg gestellt. Vom Äthersystem aus erscheinen die bewegten Ladungen als zwei antiparallele Ströme. Diese sollten sich abstoßen, was zu einem Drehmoment proportional zu $(v/c)^2$ führen müßte, aber nicht beobachtet wird.

Versuch von Sagnac. Bei diesem Experiment wird ein Lichtstrahl an einem halbdurchlässigen Spiegel in zwei Teile zerlegt, die dann gegensinnig einen durch Spiegel abgesteckten, geschlossenen Weg durchlaufen und am Ende zur Interferenz gebracht werden. Lichtquelle, Spiegel und Beobachtungsgerät stehen auf einer Scheibe, die man rotieren läßt. Dabei tritt eine Streifenverschiebung entsprechend einer Laufzeitdifferenz der beiden Strahlen von $\Delta t = 4\omega F/c^2$ auf (ω = Kreisfrequenz, F = vom Lichtweg berandete Fläche). Bedeutungsvoll wurde dieses Experiment in Gestalt des Versuches von Michelson und Gale zum optischen Nachweis der Erdrotation.

Geschwindigkeitsabhängigkeit der Masse. Mißt man die Ablenkung elektrischer Teilchen in elektromagnetischen Feldern, so stimmen die Resultate mit der Lex secunda überein, wenn Impuls und Masse durch $p = m(v) \cdot v$; $m(v) = m_0 / \sqrt{1 - v^2/c^2}$ gegeben sind. Diese Geschwindigkeitsabhängigkeit der Masse ist mit großer Präzision experimentell geprüft worden.

3.1.1.4. Deutung der Experimente. Die Experimente von MICHELSON, TROUTON/NOBLE, der quadratische Doppler-Effekt und die Geschwindigkeitsabhängigkeit der Masse zeigten die Unhaltbarkeit der nichtrelativistischen Theorie bei Verwendung eines ruhenden Äthers. Zu ihrer Rettung postulierte man eine Mitführung desselben durch die Bewegung des Mediums. Widerlegt wurde diese Hypothese durch die Experimente von SAGNAC und MICHELSON/GALE, die zeigten, daß die Erdrotation den Äther nicht mitführt. Damit war die Konzeption eines ausgezeichneten Bezugssystems, des Äthersystems, gescheitert.

3.1.2. Grundlagen der Speziellen Relativitätstheorie

3.1.2.1. Relativitätsprinzip. Die angedeuteten Schwierigkeiten konnten 1905 von EINSTEIN durch sein *Spezielles Relativitätsprinzip* überwunden werden. „Durch keinerlei physikalische Verfahren kann man ein ausgezeichnetes Inertialsystem finden." Im *Prinzip der Konstanz der Lichtgeschwindigkeit* fordert man zusätzlich, daß die Ausbreitungsgeschwindigkeit des Lichtes im Vakuum in jedem Inertialsystem in jeder Richtung gleich groß ist.

3.1.2.2. Lorentz-Transformationen. Die Untersuchung der Lichtausbreitung in zwei Inertialsystemen mit Hilfe der beiden Postulate 2.1. führt eindeutig auf die Lorentz-Transformationen, die anstelle der nicht korrekten Galilei-Transformationen beim Übergang zwischen Inertialsystemen verwendet werden müssen. Sie lauten

$$x^{i'} = a^{i'}_k x^k + d^{i'}, \quad a^{i'}_k a^{j'}_l \eta^{kl} = \eta^{i'j'}.$$

$a^{i'}_k, d^{i'} = \text{konst.}$; $(x^1 = x, x^2 = y, x^3 = z)$ sind kartesische Koordinaten, $x^4 = ct$.

Spezielle eigentliche Lorentz-Transformationen. Sind die Achsen der beiden Inertialsysteme gleichgerichtet, und zeigt die Relativgeschwindigkeit in Richtung der x-Achse, so erhält man die speziellen eigentlichen Lorentz-Transformationen.

$$x' = \frac{x - vt}{\sqrt{1 - v^2/c^2}}, \quad t' = \frac{t - vx/c^2}{\sqrt{1 - v^2/c^2}}, \quad y' = y, \quad z' = z$$

Allgemeine eigentliche Lorentz-Transformationen. So nennt man Transformationen zwischen Inertialsystemen, deren räumliche Achsen gegeneinander verdreht sind (ohne Spiegelung) und deren Relativgeschwindigkeit beliebige Richtung hat. Sie lauten

$$t' = \frac{t - v_l x^l/c^2}{\sqrt{1 - v^2/c^2}},$$

$$x^{c'} = d^{c'}_d \left\{ x^d + v^d \left[\frac{v_l x^l}{v^2} \left(\frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right) - \frac{t}{\sqrt{1 - v^2/c^2}} \right] \right\}$$

$$\text{mit } v^2 = v_l v^l, \quad v_l = \eta_{ld} v^d, \quad \eta_{ld} = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & & -1 \end{pmatrix}.$$

Sie enthalten mit den drei Komponenten der Relativgeschwindigkeit und den drei freien Richtungskosinus $d^{c'}_d$ sechs freie Parameter.

Uneigentliche Lorentz-Transformationen. Sie setzen sich aus einer eigentlichen Lorentz-Transformation und einer der folgenden Spiegelungen zusammen:

Raum-Zeit-Spiegelung	$t' = -t, \; x' = -x, \; y' = -y, \; z' = -z$
Raum-Spiegelung	$t' = t, \; x' = -x, \; y' = -y, \; z' = -z$
Zeit-Spiegelung	$t' = -t, \; x' = x, \; y' = y, \; z' = z$

Diese Spiegelungen spielen in der Elementarteilchentheorie eine Rolle.

Lorentz-Gruppe. Die Lorentz-Transformationen bilden eine Gruppe mit sechs Parametern. Nimmt man die vier Translationsfreiheitsgrade (Verschiebungen des Koordinatenursprungs) hinzu, kommt man zur zehnparametrigen Poincaré-Gruppe.

3.1.2.3. Kinematische Konsequenzen der Lorentz-Transformationen. Die Lorentz-Transformationen haben bezüglich Raum, Zeit und Gleichzeitigkeit weitreichende Konsequenzen, die experimentell gut bestätigt sind. Es geht dabei immer um Längen- und Zeitmessungen zweier Beobachter in den Inertialsystemen K und K'. Sie benutzen dazu Einheitsmaßstäbe und Normaluhren, die in *ihrem* Inertialsystem ruhen.

Relativierung der Gleichzeitigkeit. Die Aussage „Zwei Ereignisse finden an den Orten A und B gleichzeitig statt" ist relativ, d. h. vom Beobachter abhängig. Registriert ein Beobachter im System K zwei Ereignisse an den Orten $A(+x_A, 0, 0)$ und $B(x_B, 0, 0)$ gleichzeitig $(t_A = t_B)$, so registriert ein Beobachter im Inertialsystem K' diese Ereignisse mit der Zeitdifferenz $v(x_B - x_A)/c^2 \sqrt{1 - v^2/c^2} = t'_A - t'_B$, also keineswegs gleichzeitig.

Längenkontraktion. Im System K' befinde sich auf der x-Achse ein ruhender Maßstab mit der „Ruhlänge" $l_0 = x'_B - x'_A$. Im System K mißt ein Beobachter die Länge des Stabes, indem er die Koordinaten der Stabenden *gleichzeitig* feststellt $(t_A = t_B)$. Er erhält $l = x_A - x_B = l_0 \sqrt{1 - v^2/c^2} < l_0$. Die Länge des Stabes hängt vom Inertialsystem ab, in dem sie gemessen wird. Ein bewegter Stab wird um den Faktor $\sqrt{1 - v^2/c^2}$ in Bewegungsrichtung verkürzt.

Zeitdilatation. Ist allgemein $T_0 = t_B - t_A$ die Dauer irgendeines Vorganges, der im Inertialsystem K an einem festen Ort abläuft, und $T = t'_B - t'_A$ die Dauer desselben Vorganges, beurteilt vom System K', so gilt $T = T_0 / \sqrt{1 - v^2/c^2} > T_0$. Demnach geht jede Uhr, die sich im Inertialsystem K bewegt, gegenüber den ruhenden Uhren in K nach. Jede Uhr geht am schnellsten in ihrem Ruhsystem.

Einsteinsches Additionstheorem der Geschwindigkeiten. Aus den Lorentz-Transformationen erhält man das Einsteinsche Additionstheorem der Geschwindigkeiten $v = (v_1 + v_2)/(1 + v_1 v_2/c^2)$. Die Addition zweier Geschwindigkeiten $v_1 < c$, $v_2 < c$ führt immer auf eine resultierende Geschwindigkeit, die kleiner als die Lichtgeschwindigkeit ist. Es ist nicht möglich, einen Vorgang auszulösen, der zur Übertragung von Informationen mit Überlichtgeschwindigkeit dienen kann.

3.1.3. Vierdimensionale Formulierung im Minkowski-Raum

3.1.3.1. Deutung der Lorentz-Transformationen als Drehung im vierdimensionalen pseudoeuklidischen Raum-

Zeit-Kontinuum („Minkowski-Raum"). Raum-Zeit-Kontinuum, Weltpunkt, Weltlinie. Das vierdimensionale Raum-Zeit-Kontinuum („Raum-Zeit") besteht aus Ereignissen (Weltpunkten). Jeder Weltpunkt wird durch vier Koordinaten $x^a (a = 1 \dots 4)$ festgelegt, von denen die ersten drei den Ort und die vierte den Zeitpunkt des Ereignisses bestimmen. Eine Weltlinie $x^a = x^a(x^4)$ beschreibt die Lebensgeschichte eines Massenpunktes oder Lichtsignals. Bewegen sich diese längs der x-Achse, so können ihre Weltlinien in der (x, x^4)-Ebene dargestellt werden (Minkowski-Diagramm, Abb. 3.1.3–1).

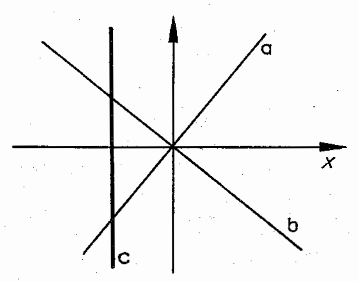


Abb. 3.1.3–1 Weltlinie für ein Lichtsignal $x = +ct$ (a), $x = -ct$ (b) und einen ruhenden Massenpunkt (c)

Pseudoeuklidischer Raum. In Analogie zum dreidimensionalen euklidischen Raum kann auch für das vierdimensionale Raum-Zeit-Kontinuum ein Abstandsbegriff, das Raum-Zeit-Intervall, eingeführt werden. Liegen die beiden Ereignisse $E_1(x^a)$ und $E_2(x^a + dx^a)$ infinitesimal benachbart, so ist es gegeben durch $ds^2 = dx^2 + dy^2 + dz^2 - (dx^4)^2 = \eta_{ab} dx^a dx^b$. $\eta_{11} = \eta_{22} = \eta_{33} = -\eta_{44} = 1$, $\eta_{ab} = 0$ sonst.

Über gleiche Indizes wird stets summiert (Einsteinsche Summenkonvention). ds^2 wird metrische Fundamentalform des vierdimensionalen pseudoeuklidischen Raumes (Minkowski-Raumes) genannt. ds^2 ist invariant unter Lorentz-Transformationen. Die Geometrie im Minkowski-Raum unterscheidet sich wegen des Minuszeichens vor dem Quadrat der Zeitkoordinate in wesentlichen Punkten von der euklidischen Geometrie, man nennt sie pseudoeuklidische Geometrie.

Lorentz-Transformationen als Pseudodrehungen im Minkowski-Raum. Die speziellen Lorentz-Transformationen können in der Form geschrieben werden: $x' = x\cosh\varphi + x^4\sinh\varphi$, $x^{4'} = x\sinh\varphi + x^4\cosh\varphi$, $\tanh\varphi = -v/c$, $x^4 = ct$. Sie sind das Analogon zu den Drehungen der Koordinatenachsen in der euklidischen (x, y)-Ebene. Im Unterschied zu diesen lassen aber diese „Pseudodrehungen" im Minkowski-Raum nicht den Winkel zwischen den Koordinatenachsen, sondern die Winkelhalbierende $x = \pm ct$. fest.

3.1.3.2. Elementare Eigenschaften des Minkowski-Raumes. Lichtkegel, zeitartige, raumartige und lichtartige Raum-Zeit-Intervalle. Weltpunkte E, die die Beziehung erfüllen $\Delta s^2 = (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - (\Delta x^4)^2 = 0$, liegen im Minkowski-Raum auf dem Mantel eines vierdimensionalen Doppelkegels mit der Spitze im Koordinatenursprung, der Lichtkegel genannt wird. Die Gleichung des Lichtkegels ist lorentzinvariant, er wird immer von denselben Raum-Zeit-Punkten gebildet. Ereignisse, die durch ein zeitartiges Intervall $((\Delta s)^2 < 0)$ getrennt sind, haben in jedem Inertialsystem die gleiche zeitliche Reihenfolge. Sind dagegen zwei Ereignisse durch ein raumartiges Intervall $((\Delta s)^2 > 0)$ getrennt, dann ist die zeitliche Reihenfolge dieser Ereignisse vom Inertialsystem abhängig. Auf der Weltlinie eines Lichtstrahles sind alle Ereignisse durch lichtartige Intervalle $(\Delta s)^2 = 0$ voneinander getrennt.

3.1.3.3. Kovarianz und geometrische Objekte. Vierervektoren, Invarianten und Tensoren im Minkowski-Raum. In direkter Analogie zu den Dreiervektoren im euklidischen Raum kann man im Minkowsi-Raum Vierervektoren einführen. Vier Zahlen (Komponenten) bilden einen Vierervektor $(U^a) = \boldsymbol{U}$, wenn sie sich unter Lorentz-Transformationen wie folgt transformieren

$$U^{a'} = a^{a'}_b U^b, \quad \text{z. B.}$$

$$a^{a'}_b = \begin{pmatrix} (1 - v^2/c^2)^{-1/2} & 0 & 0 & -v/c(1 - v^2/c^2)^{-1/2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -v/c(1 - v^2/c^2)^{-1/2} & 0 & 0 & (1 - v^2/c^2)^{-1/2} \end{pmatrix}.$$

Aus Vierervektoren können Invarianten gebildet werden, das sind Größen, die in jedem Inertialsystem denselben Zahlenwert haben. Das Skalarprodukt $\eta_{ab} U^a V^b$ zweier Vierervektoren ist ein Beispiel. Gilt $\eta_{ab} U^a U^b \gtreqless 0$, nennt man den Vierervektor raumartig, zeitartig bzw. lichtartig. Unter einem (kontravarianten) Tensor n-ter Stufe versteht man die 4^n-Zahlen (Komponenten) $t^{a_1 \dots a_n}$, die sich bei Lorentz-Transformationen nach dem Schema umrechnen:

$$t^{b'_1 \dots b'_n} = a^{b'_1}_{c_1} \dots a^{b'_n}_{c_n} t^{c_1 \dots c_n}.$$

Ein Tensor n-ter Stufe ist symmetrisch bzw. antisymmetrisch bezüglich der Indizes b_r und b_s, wenn gilt

$$t^{b_1 \dots b_r \dots b_s \dots b_n} = \pm t^{b_1 \dots b_s \dots b_r \dots b_n}.$$

$\eta_{ab} = \eta_{ba}$ ist ein spezieller symmetrischer Tensor 2. Stufe. Zwischen den ko- und kontravarianten Komponenten dieses metrischen Tensors bestehen die Relationen $\eta^{ab}\eta_{bc} = \eta^a_b$; $\eta_{11} = \eta_{22} = \eta_{33} = -\eta_{44} = 1$, $\eta_{ab} = 0$ sonst. Will man im Ausdruck $\eta_{ab} U^a V^b =$ Inv. den metrischen Tensor sparen, so muß man durch $U_a = \eta_{ab} U^b$ die kovarianten Komponenten U_a des Vierervektors $\boldsymbol{U}$ einführen. Die kovarianten und die gemischten Komponenten eines kontravarianten Tensors t^{ab} erhält man analog durch „Index ziehen" $t_{ab} = \eta_{ad}\eta_{be} t^{de}$, $t^b_a = \eta_{ac} t^{cb}$; $t^a_b = \eta_{bc} t^{ac}$. Addition und Subtraktion zweier Tensoren gleicher Stufe, die sich auf denselben Weltpunkt beziehen, liefern wieder einen Tensor. Für die Tensoraddition gelten das Assoziativ- und das Kommutativgesetz. Das äußere Produkt zweier Tensoren der Stufe r bzw. s, die sich auf denselben Weltpunkt beziehen müssen, liefert einen Tensor der Stufe $r + s$.

$$t^{a_1 \dots a_r b_1 \dots b_s}(x^a) = \underset{\mathrm{I}}{t^{a_1 \dots a_r}(x^a)} \cdot \underset{\mathrm{II}}{t^{b_1 \dots b_s}(x^a)}.$$

Vierervektoren, Tensoren und Spinoren sind die Darstellungen der Lorentz-Gruppe.

Formulierung von Naturgesetzen und spezielles Relativitätsprinzip. Tensorgleichungen haben in allen Inertialsystemen dieselbe Form. Das spezielle Relativitätsprinzip verlangt dann, daß die Naturgesetze als Tensor- oder Spinorgleichungen schreibbar sein müssen.

Eigenzeit und Vierergeschwindigkeit. Das Eigenzeitintervall $d\tau$ ist mit dem invarianten Raum-Zeit-Intervall durch $d\tau = \sqrt{-1/c^2 ds^2} = dt\sqrt{1 - \dot{r}^2/c^2}$ verknüpft und daher selbst eine Invariante. Eine mit dem Massenpunkt verbundene, d. h. mitbewegte Normal-Uhr zeigt genau dessen Ei-

genzeit an. Das Integral über das Eigenzeitintervall $\Delta\tau$ $= \int_A^B \sqrt{1-\dot{r}^2/c^2}\,dt$ ist wegabhängig. An Stelle der nichtinvarianten Dreiergeschwindigkeit $v = dr/dt$ der Newtonschen Mechanik tritt die Vierergeschwindigkeit $U^a = dx^a/d\tau$. Weil dx^a ein Vierervektor und $d\tau$ eine Invariante ist, hat diese Definition in jedem Inertialsystem dieselbe Form. Die Vierergeschwindigkeit ist ein zeitartiger Vektor mit festem Betrag. $\eta_{ab} U^a U^b = -c^2$. Sie ist Tangentenvektor an eine zeitartige Teilchenweltlinie. Ihre Komponenten lauten $U^a = dx^a/d\tau = (v/\sqrt{1-v^2/c^2}\,,\ c/\sqrt{1-v^2/c^2})$.

Levi-Civitascher Pseudotensor. Die Komponenten des vollständig antisymmetrischen Einheitspseudotensors ε^{abcd} haben in allen Inertialsystemen denselben Wert. Vertauscht man in ε^{abcd} zwei beliebige Indizes, so ändert sich das Vorzeichen. $\varepsilon^{abcd} = -\varepsilon^{bacd} = -\varepsilon^{abdc} = -\varepsilon^{cbad}$. Daraus folgt, daß alle Komponenten ε^{abcd} mit zwei gleichen Indizes verschwinden. Dasselbe gilt auch für die kovarianten Komponenten des Pseudotensors, die man wie üblich durch Indexziehen erhält $\varepsilon_{abcd} = \eta_{ae}\eta_{bf}\eta_{cg}\eta_{dh}\varepsilon^{efgh}$. Die von Null verschiedenen Komponenten des Pseudotensors ergeben sich aus $\varepsilon^{1234} = +1$ bzw. $\varepsilon_{1234} = -1$ durch Vertauschen der Indizes. Das Transformationsgesetz des Pseudotensors lautet $\varepsilon^{a'b'c'd'} = a_e^{a'} a_f^{b'} a_g^{c'} a_h^{d'} \varepsilon^{efgh} \operatorname{sign}|a_k^{i'}|$.

3.1.4. Relativistische Mechanik

3.1.4.1. Grundgesetz der Mechanik. Das Grundgesetz der speziell-relativistischen Mechanik lautet $dp^a/d\tau = d/d\tau(m_0 U^a) = F^a$. Dabei ist $p^a = m_0 U^a$ der Viererimpuls, U^a die Vierergeschwindigkeit, F^a die Viererkraft und m_0 eine skalare Invariante, die Ruhmasse. Für kleine Geschwindigkeiten muß das neue Grundgesetz in die alte Lex secunda übergehen. Rechnet man die Eigenzeit auf die Koordinatenzeit um, so erhält man

$$\frac{d}{dt}\left(\frac{m_0 v^\mu}{\sqrt{1-v^2/c^2}}\right) = F^\mu\sqrt{1-v^2/c^2} = K^\mu;$$

$$\frac{d}{dt}\left(\frac{m_0 c}{\sqrt{1-v^2/c^2}}\right) = F^4\sqrt{1-v^2/c^2}\,,$$

$\mu = 1\ldots3$. Es gilt $F^a U_a = 0$, d.h. $F^4 = (1/c)F_\mu\,dx^\mu/dt$, $\mu = 1\ldots3$. Wenn Masse und Dreierkraft durch $m = m_0/\sqrt{1-v^2/c^2}$ bzw. $K = e_\mu F^\mu\sqrt{1-v^2/c^2}$ definiert werden, erhält man das Grundgesetz der Newtonschen Mechanik. Die Masse wird offensichtlich geschwindigkeitsabhängig.

3.1.4.2. Leistungsbilanz und Energiesatz. Ist K im betrachteten Koordinatensystem aus einem Potential ableitbar, $K = -\operatorname{grad} V$, dann wird aus $d(mc^2)/dt = F^4\sqrt{1-v^2/c^2} = Kv$ der allerdings nicht invariante „Energiesatz" $mc^2 + V(r) = m_0c^2/\sqrt{1-v^2/c^2} + V(r) = W_0 =$ konst. Man bezeichnet die Größe mc^2 als „Energie des Teilchens" $E = mc^2 = m_0c^2/\sqrt{1-v^2/c^2} = m_0c^2 + (m_0/2)v^2 + \ldots$ Genähert hat E die Gestalt kinetische Energie plus Konstante. Der Anteil m_0c^2, die Ruhenergie des Teilchens, ist stets vorhanden und Bestandteil des Energiesatzes. Ruhmassenenergie kann auch in andere Energieformen verwandelt werden, z.B. in Bewegungsenergie bei Teilchenumwandlungen. E und m sind keine Invarianten, sondern entsprechen der vierten Komponente des Impulsvierervektors $p^a = (P, E/c)$. Die neue „Lex secunda" mit $m = m(v)$ und $E = mc^2$ wurde durch zahlreiche Experimente nachgeprüft. Es zeigte sich ausgezeichnete Übereinstimmung mit dem Experiment, z.B. beim Stoß von Teilchen in der Kernphysik oder bei Ablenkung elektrisch geladener Teilchen in elektromagnetischen Feldern.

3.1.5. Elektrodynamik

3.1.5.1. Vierdimensionale Formulierung und Maxwell-Gleichungen. Im Gegensatz zur Mechanik ist die Elektrodynamik bereits eine lorentzinvariante Theorie. Allerdings kann man die Dreiervektoren der elektrischen und magnetischen Feldstärken nicht als invariante Elemente verwenden. Sie lassen sich allein (ohne Zuhilfenahme anderer Größen) auch nicht als Vierervektoren schreiben. Sie sind vielmehr Komponenten vierdimensionaler antisymmetrischer Tensoren, nämlich des *Feldstärketensors* B_{ik} und des *Erregungstensors* H_{ik}.

$$(B_{ik}) = \begin{pmatrix} 0 & +B_z & -B_y & E_x/c \\ -B_z & 0 & +B_x & E_y/c \\ +B_y & -B_x & 0 & E_z/c \\ -E_x/c & -E_y/c & -E_z/c & 0 \end{pmatrix},$$

$$(H^{ik}) = \begin{pmatrix} 0 & H_z & -H_y & -cD_x \\ -H_z & 0 & H_x & -cD_y \\ H_y & -H_x & 0 & -cD_z \\ cD_x & cD_y & cD_z & 0 \end{pmatrix}.$$

B_α, E_α, H_α und D_α sind die kartesischen Komponenten der magnetischen Induktion, der elektrischen und magnetischen Feldstärke und der dielektrischen Verschiebung. $c = 1/\sqrt{\varepsilon_0\mu_0}$. Mit dem aus Strom- und Ladungsdichte aufgebauten *Stromdichtevektor* j^i, $(j^i) = (j, \varrho c)$ haben die *Maxwell-Gleichungen* (in jedem Minkowsky-System) die Gestalt:

(1) $B_{\langle ij,k\rangle} = 0$, (2) $H^{ik}{}_{,k} = j^i$.

Dabei wird die partielle Ableitung durch Komma dargestellt, und der Operator $\langle ijk\rangle$ bedeutet zyklische Vertauschung der Indizes:

$$B_{ik,l} \equiv \partial B_{ik}/\partial x^l;\ B_{\langle ij,k\rangle} \equiv B_{ij,k} + B_{jk,i} + B_{ki,j}$$
$$\equiv \partial B_{ij}/\partial x^k + \partial B_{jk}/\partial x^i + \partial B_{ki}/\partial x^j.$$

Weiter gilt die „Einsteinsche Summenkonvention": Gleiche Indizes werden automatisch über den entsprechenden Wertevorrat summiert:

$$H^{1k}{}_{,k} \equiv H^{11}{}_{,1} + H^{12}{}_{,2} + H^{13}{}_{,3} + H^{14}{}_{,4}$$
$$= \partial H_z/\partial y - \partial H_y/\partial z - \partial cD_x/\partial ct = j_x.$$

In *Dreierform* entstehen aus (1) und (2) die Maxwell-Gleichungen in üblicher Gestalt, und zwar ergeben jeweils:

$$B_{\langle ij,k\rangle} = 0 \begin{cases} \nearrow \operatorname{div} B = 0 \\ \searrow \operatorname{rot} E = -\dot{B}; \end{cases} \qquad H^{ik}{}_{,k} = j^i \begin{cases} \nearrow \operatorname{div} D = \varrho \\ \searrow \operatorname{rot} H = j + \dot{D} \end{cases}$$

Materialgleichung. Den Zusammenhang zwischen B_{ik} und H_{ik} kann man in der Form darstellen.

(3) $B_{ik} = \mu_0 H_{ik} + M_{ik}$.

$$(M_{ik}) = \begin{pmatrix} 0 & M_z & -M_y & -P_x/(\varepsilon_0 c) \\ -M_z & 0 & M_x & -P_y/(\varepsilon_0 c) \\ M_y & -M_x & 0 & -P_z/(\varepsilon_0 c) \\ P_x/(\varepsilon_0 c) & P_y/(\varepsilon_0 c) & P_z/(\varepsilon_0 c) & 0 \end{pmatrix}.$$

Die Polarisation $\boldsymbol{P}$ und die Magnetisierung $\boldsymbol{M}$ hängen in üblicher Weise von $\boldsymbol{E}$ bzw. $\boldsymbol{H}$ ab. ($H_{\alpha 4} = -H^{\alpha 4}$, $c^2 = 1/[\varepsilon_0 \mu_0]$). Ohmsches Gesetz vgl. 3.1.5.7.
Den Anschluß an die Mechanik (vgl. 3.1.4.1) erhält man über die *Viererkraft auf eine Ladung q im elektromagnetischen Feld:* $F_i = q B_{ik} u^k$.

$$(F_i) = \frac{(\boldsymbol{K}; \boldsymbol{K}\cdot \boldsymbol{v}/c)}{\sqrt{1-(v/c)^2}} \to \boldsymbol{K} = q\{\boldsymbol{E} + [\boldsymbol{v}\times \boldsymbol{B}]\}.$$

3.1.5.2. Duale Feldtensoren. Durch die Definitionen $\tilde{B}^{ik} = (1/2)\,\varepsilon^{ikrs} B_{rs}$, $\tilde{H}_{ik} = (1/2)\,\varepsilon_{ikrs} H^{rs}$ werden der duale Feldstärketensor und der duale Erregungstensor definiert.

$$(\tilde{B}^{ik}) = \begin{pmatrix} 0 & +\frac{E_z}{c} & -\frac{E_y}{c} & B_x \\ -\frac{E_z}{c} & 0 & +\frac{E_x}{c} & B_y \\ +\frac{E_y}{c} & -\frac{E_x}{c} & 0 & B_z \\ -B_x & -B_y & -B_z & 0 \end{pmatrix},$$

$$(\tilde{H}_{ik}) = \begin{pmatrix} 0 & cD_z & -cD_y & -H_x \\ -cD_z & 0 & cD_x & -H_y \\ +cD_y & -cD_x & 0 & -H_z \\ H_x & H_y & H_z & 0 \end{pmatrix}.$$

Ihr Nutzen liegt erstens in den „dualen" Feldgleichungen (die manchmal günstiger sind)

$$\tilde{B}^{ik}{}_{,k} = 0 \qquad \leftrightarrow B_{(ik,e)} = 0,$$
$$\tilde{H}_{(ij,k)} = -\varepsilon_{ijkl} j^l \qquad \leftrightarrow H^{ik}{}_{,k} = j^i,$$

und zweitens im Plätzetausch der elektrischen und magnetischen Komponenten. Sie können so z. B. wechselseitig aus der 4. Spalte herausprojiziert werden.

3.1.5.3. Transformation des Feldstärketensors. Daß $\boldsymbol{E}$ und $\boldsymbol{B}$ bzw. $\boldsymbol{H}$ und $\boldsymbol{D}$ zu Tensoren zusammengefaßt werden müssen, sieht man nicht nur in der dadurch möglichen vierdimensionalen kovarianten Form der Maxwell-Gleichungen, sondern noch deutlicher im Transformationsverhalten bei Lorentz-Transformation zwischen zwei I-Systemen. Bewegt sich das System $\{x^{i'}\}$ gegen $\{x^i\}$ mit der Geschwindigkeit v in x-Richtung, gelten die speziellen eigentlichen Lorentz-Transformationen:

$$E'_x = E_x,\; E'_y = \frac{E_y - (v/c)\,B_z}{\sqrt{1-(v/c)^2}},\; E'_z = \frac{E_z + (v/c)\,B_y}{\sqrt{1-(v/c)^2}},$$
$$B'_x = B_x,\; B'_y = \frac{B_y + (v/c)\,E_z}{\sqrt{1-(v/c)^2}},\; B'_z = \frac{B_z - (v/c)\,E_y}{\sqrt{1-(v/c)^2}},$$

was experimentell nachprüfbar ist. Die $\boldsymbol{E}$- und $\boldsymbol{B}$-Felder werden gemischt. Aus einem reinen elektrischen oder magnetischen Feld entsteht in einem dagegen bewegten System ein zusätzliches magnetisches bzw. elektrisches Feld. Beispielsweise zeigt ein in einem Magnetfeld bewegtes Atom – vermöge des im Ruhsystem des Atoms erzeugten elektrischen Feldes – eine Stark-Aufspaltung entsprechender Linien.

3.1.5.4. Transformation einer ebenen Welle, Aberration und Doppler-Effekt. Eine elektromagnetische Welle

$$\boldsymbol{E} = \boldsymbol{E}_0 \cos(\boldsymbol{kr} - \omega t - \delta) \equiv \boldsymbol{E}_0 \cos(k_s x^s - \delta) \equiv \operatorname{Re}\{\hat{\boldsymbol{E}}_0 \mathrm{e}^{\mathrm{i}k_s x^s}\},$$
$$\boldsymbol{H} = \boldsymbol{H}_0 \cos(\boldsymbol{kr} - \omega t - \delta) \equiv \boldsymbol{H}_0 \cos(k_s x^s - \delta) \equiv \operatorname{Re}\{\hat{\boldsymbol{H}}_0 \mathrm{e}^{\mathrm{i}k_s x^s}\}$$

transformiert sich in obiger Weise. Dabei bleibt die Phase $\Phi = k_s x^s \equiv \boldsymbol{kr} - \omega t$ invariant. Daraus folgen die Transformationsformeln für den Wellenzahlvierervektor

$$(k_s) = (\boldsymbol{k}, -\omega/c) = (\boldsymbol{n}\cdot 2\pi/\lambda, -2\pi\nu/c) = \omega/c\cdot(\boldsymbol{n}c/c_{ph}, -1);$$
$$k^{1'} = \frac{k^1 - (v/c)\,k^4}{\sqrt{1-(v/c)^2}},\quad k^{2'} = k^2,\quad k^{3'} = k^3,$$
$$k^{4'} = \frac{k^4 - (v/c)\,k^1}{\sqrt{1-(v/c)^2}}.$$
$$c^2 = 1/(\varepsilon_0\mu_0),\quad c_{ph}^2 = 1/(\varepsilon\mu),\quad c_{ph} = \nu\lambda.$$

Aus der Richtungsänderung von $\boldsymbol{k}$ bzw. $\boldsymbol{n}$ erhält man die *Aberration*, die Änderung der Richtung des Lichtstrahles ($k_1 = |\boldsymbol{k}|\cos\vartheta$), während die 4. Beziehung den Doppler-Effekt liefert:

$$\cos\vartheta' = [\cos\vartheta - vc_{ph}/c^2]/(1 - v/c\cos\vartheta),$$
$$\nu' = \nu\frac{1-(v/c_{ph})\cos\vartheta}{\sqrt{1-(v/c)^2}}$$

Ist $\{x^{i'}\}$ das Ruhsystem einer Lichtquelle ($\nu' = \nu_0$), gilt:

$$\nu = \nu_0 \frac{\sqrt{1-(v/c)^2}}{1-(v/c_{ph})\cdot\cos\vartheta}.$$

Für kleine Geschwindigkeiten (bei Entwicklung bis v/c) ergeben sich die nichtrelativistischen Werte. Relativistische Abweichungen treten erst bei den Gliedern ab $(v/c)^2$ auf: „quadratischer Doppler-Effekt". Seine Messung bestätigt die Aussagen der Speziellen Relativitätstheorie.

3.1.5.5. Viererpotential und Eichung. Hinreichend und notwendig zur Erfüllung der Maxwell-Gleichung (1) ist die Existenz eines Viererpotentials A^i

$$\left.\begin{matrix} B_{(ij,k)} = 0 \\ (\text{bzw. } \tilde{B}^{ij}{}_{,j} = 0) \end{matrix}\right\} \leftrightarrow \left\{\begin{matrix} B_{ij} = A_{j,i} - A_{i,j} & (A^i) = (\boldsymbol{A}, U/c) \\ (\tilde{B}^{ij} = \varepsilon^{ijkl} A_{l,k}) & \boldsymbol{B} = \operatorname{rot}\boldsymbol{A}; \end{matrix}\right.$$
$$\boldsymbol{E} = -\partial\boldsymbol{A}/\partial t - \nabla U.$$

Eichung. Meßbar sind nur die Feldstärken und Erregungen. Zu einem gegebenen B_{ik} gehört eine ganze Klasse von Viererpotentialen, neben A_i genau auch $A_i = A_i + \chi_{,i}$ mit beliebigem $\chi(\boldsymbol{r}, t)$. Eine Eichung ist eine spezielle Festlegung von A_i bzw. χ (eventuell zur Vereinfachung von Differentialgleichungen, wie z. B. die Lorentz-Konvention).
Lorentz-Konvention: $A^i{}_{,i} = 0$; $\nabla\boldsymbol{A} + \partial(U/c)/\partial ct = 0$.

3.1.5.6. Elektrodynamik im Vakuum. Im Vakuum gilt $M_{ik} = 0$ und nach (3): $B_{ik} = \mu_0 H_{ik}$. Damit schreiben sich die Maxwell-Gleichungen (1) und (2):

(1') $B_{(ij,k)} = 0$; $B_{ij} = A_{j,i} - A_{i,j}$. Eichung $A^i{}_{,i} = 0$.
(2') $B^{ik}{}_{,k} = \mu_0 j^i$; $A^{k,i}{}_{,k} - A^{i,k}{}_{,k} \equiv -\eta^{kl} A^i{}_{,k,l} \equiv -\Box A^i = \mu_0 j^i$

Für A^i entsteht daraus die inhomogene Wellengleichung mit Lorentz-Konvention:

$$\left.\begin{matrix} \Box A^i = -\mu_0 j^i, \\ A^i{}_{,i} = 0 \end{matrix}\right\} \leftrightarrow \left\{\begin{matrix} \Box\boldsymbol{A} = -\mu_0\boldsymbol{j},\ \Box U = -\varrho/\varepsilon_0, \\ \nabla\boldsymbol{A} + (1/c^2)\,\partial U/\partial t = 0. \end{matrix}\right.$$

Die physikalisch relevanten Lösungen haben die Gestalt *retardierter Potentiale.* Von einem atomistischen Standpunkt aus betrachtet kann man (wenn man Atome, Ionen usw. als Ladungspunkte und Ströme ansieht) auch die Elektrodynamik in Medien als Elektrodynamik im Vakuum auffassen („Elektronentheorie").

3.1.5.7. Elektrodynamik in Medien. Durch (1), (2) und (3) wird die Elektrodynamik in Medien dargestellt. Als Besonderheit existiert aber hier als zusätzlicher Vektor die Vierer-

geschwindigkeit $u^i(x)$ für jeden Punkt des Mediums. Dadurch kann man kovariant vierdimensionale Vektoren der elektrischen und magnetischen Feldstärke, der dielektrischen Verschiebung und der magnetischen Induktion bilden

$$\hat{E}^i = B_{ik} u^k, \quad \hat{H}_i = (1/2)\,\varepsilon_{iklm} H^{lk} u^m/c = -\tilde{H}_{in} u^n/c,$$
$$\hat{D}^i = H^{ik} u_k/c^2, \quad \hat{B}^i = (1/2)\,\varepsilon^{iklm} B_{lk} u_m/c = -\tilde{B}^{in} u_n/c.$$

Die eigentliche physikalische Bedeutung haben diese Vektoren im lokalen Ruhsystem des Mediums, wo $u^i = c\,\delta^i_4$ ist. Benutzen kann man diese Vektoren, um die Materialgleichungen multiplikativ zu formulieren:

$$\hat{D}_i = \varepsilon \hat{E}_i \rightarrow H_{ik} u^k/c^2 = \varepsilon B_{ik} u^k$$
$$\hat{B}_i = \mu \hat{H}_i \rightarrow \tilde{B}_{ik} u^k/c = \mu \tilde{H}_{ik} u^k/c.$$

Analog lassen sich diese Beziehungen auch für anisotrope Medien verwenden: $\hat{D}_i = \varepsilon_{ik} \hat{E}^k$, $\hat{B}_i = \mu_{ik} \hat{H}^k$, sowie für das *Ohmsche Gesetz*: $j^i = j^i_{\text{conv}} + j^i_{\text{cond}}$.

$$j^i_{\text{conv}} = \varrho_0 u^i, \qquad j^i_{\text{cond}} = \sigma \hat{E}^i = \sigma B^{ik} u_k.$$
$$\boldsymbol{j}_{\text{conv}} = \frac{\varrho_0 \boldsymbol{v}}{\sqrt{1-(v/c)^2}}, \qquad \boldsymbol{j}_{\text{cond}} \equiv \boldsymbol{e}_\alpha j^\alpha = \frac{\sigma\{\boldsymbol{E} + [\boldsymbol{v} \times \boldsymbol{B}]\}}{\sqrt{1-(v/c)^2}}.$$

Dabei sind σ die Leitfähigkeit und ϱ_0 die Ladungsdichte im Ruhsystem.

3.1.5.8. Feldinvarianten. Aus dem Feldstärketensor (aus $\boldsymbol{E}$ und $\boldsymbol{B}$) lassen sich als Invariante (bzw. Pseudoinvariante) nur zwei Ausdrücke bilden:

$$B_{mn} B^{mn} = 2\,(\boldsymbol{B}^2 - \boldsymbol{E}^2/c^2) = \text{Invariante}$$
$$= 2\mu_0 (\boldsymbol{B}^2/\mu_0 - \varepsilon_0 \boldsymbol{E}^2).$$
$$\tilde{B}_{mn} B^{mn} = 4\boldsymbol{E}\boldsymbol{B}/c = \text{Pseudoinvariante}.$$

3.1.6. Lagrange-Formalismus

3.1.6.1. N-Teilchen im elektromagnetischen Feld. Im Vakuum lassen sich die Bewegungsgleichungen geladener Teilchen im elektromagnetischen Feld sowie die Feldgleichungen selbst aus einem gemeinsamen *Variationsprinzip* ableiten

$$\delta I = \delta(1/c) \int_{V_4} \{L_{\text{mech}} + L_{\text{elm}} + L_{\text{WW}}\}\, \mathrm{d}^4x = 0,$$
$$L_{\text{mech}} = -\mu_{(0)} c^2, \quad L_{\text{WW}} = j^k A_k = \varrho_0 u^k A_k,$$
$$L_{\text{elm}} = -1/(4\mu_0)(A_{k,i} - A_{i,k})(A_{s,r} - A_{r,s})\,\eta^{ir}\eta^{ks} = -(1/4)\,B_{ik} H^{ik}.$$

Dabei ist $\mu_{(0)}$ die Ruhmassendichte und ϱ_0 die Ladungsdichte im lokalen Ruhsystem. Variiert wird über A_i und $x^i_A(\lambda_A)$ (= Ort als kanonische Variable des Teilchens Nr. A – $A = 1 \ldots N$ – abhängig von beliebigen Parameter λ_A). I_{mech} und I_{WW} werden im lokalen Ruhsystem über den Ortsraum der (nahezu) punktförmigen Teilchen integriert:

$$I_{\text{mech}} = -\sum_A \int \mathrm{d}\tau_A\, m_{A0} c^2, \quad I_{\text{WW}} = \sum_A e_A \int A_\alpha (\mathrm{d}x^\alpha_A/\mathrm{d}\tau_A)\, \mathrm{d}\tau_A.$$

τ_A ist die Eigenzeit. Zur bequemeren Variation (ohne Nebenbedingungen) rechnet man auf beliebige Parameter λ_A um und kehrt erst am Ende wieder zu τ_A zurück. m_{A0} und e_A sind Ruhmasse und Ladung des A-ten Teilchens. Die Variation liefert:

$$\delta A_i \rightarrow H^{ik}{}_{,k} = j^i \text{ (inhomogenes Maxwell-System)}$$
$$\delta x^i_A \rightarrow m_{A0} \mathrm{d}^2 x^i_A/\mathrm{d}\tau^2_A = e_A [A_{k,i} - A_{i,k}](\mathrm{d}x^k_A/\mathrm{d}\tau_A) = e_A B_{ik} u^k_A.$$

(Bewegungsgleichung eines Massenpunktes im elektromagnetischen Feld (vgl. 3.1.4.1.). A_i ist hier unabhängige Variable, B_{ik} nur Abkürzung, für die das zyklische Maxwell-System $B_{(ik,l)} = 0$ automatisch gilt.

Dreidimensional formuliert hat das Variationsprinzip die Form

$$I_{\text{elm}} = \int \mathrm{d}t \int \mathrm{d}^3\boldsymbol{r}(1/2)\,[\varepsilon_0 \boldsymbol{E}^2 - \boldsymbol{B}^2/\mu_0],$$
$$I_{\text{mech}} + I_{\text{WW}} = \sum_A \int \mathrm{d}t \{-m_{A0} c^2 \sqrt{1-(\dot{\boldsymbol{r}}_A/c)^2} - e_A U(\boldsymbol{r}_A, t)$$
$$+ e_A \dot{\boldsymbol{r}}_A \boldsymbol{A}\}.$$

3.1.6.2. Hamilton-Formalismus. Einen solchen kann man nicht konstruieren, weil die kanonischen Impulse $\Pi^{ik} = \partial L/\partial A_{i,k} = (A^{k,i} - A^{i,k})/c$ sich nicht eindeutig nach den $A_{i,k}$ auflösen lassen.

3.1.6.3. Variationsverfahren für elektromagnetische Felder in Medien. Mit der Materialgleichung $H^{ik} = H^{ik}(B_{rs})$ ohne Hysterese: $\partial H^{ik}/\partial B_{rs} = \partial H^{rs}/\partial B_{ik}$ gilt hier

$$L = -\frac{1}{2} \int_{(0)}^{(B_{ik})} H^{ik}(B_{rs})\, \mathrm{d}B_{ik} + j^i A_i, \quad B_{ik} = A_{k,i} - A_{i,k}$$
$$\delta A_i \rightarrow H^{ik}_{,k} = j^i.$$

3.1.6.4. Energie-Impuls-Tensor. Aus der Lagrange-Dichte $L(\varphi_A; \varphi_{A,k})$ eines Variationsprinzips mit beliebigen Feldfunktionen φ_A kann man einen Energie-Impuls-Tensor (negativer kanonischer E-I-Tensor)

$$\Theta^k_i = -\partial L/\partial \varphi_{A,k} \cdot \varphi_{A,i} + \delta^k_i L \equiv -T^{(\text{can})k}_i \rightarrow \Theta^k_{i,k} = 0$$

bilden, für den ein „Erhaltungssatz" $\Theta^k_{i,k} = 0$ gilt. Diese Tensoren sind nicht symmetrisch. Durch Zufügen eines „Superpotentials" $F^{i[kl]}$ kann man neue divergenzfreie Tensoren konstruieren

$T^{ik} = \Theta^{ik} + F^{i[kl]}{}_{,l}$, $F^{i[lk]} = -F^{i[kl]} \rightarrow T^{ik}_{,k} = 0$, die sich symmetrisieren lassen. Eine solche symmetrische Form verlangt die Allgemeine Relativitätstheorie. Sie folgt auch aus dem Drehimpulserhaltungssatz.

3.1.6.5. Erhaltungssätze und Energie-Impuls-Vektor. Aus $\Theta^k_{i,k} = 0$ folgt mit Gaußschem Satz

$$(1/c)\,\mathrm{d}/\mathrm{d}t \int_{V_3} \Theta^4_i \mathrm{d}^3x + \int_{[V_3]} \Theta^\beta_i \mathrm{d}^2 f_\beta = 0,$$
$$(\mathrm{d}^2 f_1 = \mathrm{d}x^2 \mathrm{d}x^3; \ldots).$$

Verschwindet Θ^β_i auf der Berandung (z. B. im Unendlichen), oder geht es dort stärker gegen Null, als die Oberfläche zunimmt, so wird der Energie-Impuls-Vektor des Feldes

$$P_i = (1/c) \int_{V_3} \Theta^4_i \mathrm{d}^3x = \text{konst.}$$

P_i behält seinen Wert, wenn man statt über V_3 über eine beliebige raumartige Hyperfläche $\sum$ integriert oder T^k_i an Stelle von Θ^k_i verwendet.

$$P_i = \frac{1}{c} \int_{V_3} \Theta^4_i \mathrm{d}^3x = \frac{1}{c} \int_\Sigma \Theta^k_i \mathrm{d}f_k = \frac{1}{c} \int_\Sigma T^k_i \mathrm{d}f_k = \frac{1}{c} \int_{V_3} T^4_i \mathrm{d}^3x.$$

3.1.6.6. Energie-Impuls-Tensor und Vektor für elektromagnetisches Vakuumfeld ohne Ströme. Θ^k_i aus 3.1.6.1. wird mit $F^{[kl]}_i = -A_i B^{kl}/\mu_0$ symmetrisiert:

$$(T^{ik}) = \begin{pmatrix} -\sigma^{\alpha\beta}_{\text{Maxw.}} & S^\alpha/c \\ S^\beta/c & w \end{pmatrix}$$

Energiedichte $w = (\varepsilon_0 E^2 + B^2/\mu_0)/2$,
Poynting-Vektor $S = [E \times H]$,
Maxwellscher Spannungstensor $\sigma^{\alpha\beta}_{\text{Maxw.}} = \varepsilon_0 E^\alpha E^\beta + B^\alpha B^\beta/\mu_0 + 1/2\delta^{\alpha\beta}\{\varepsilon_0 E^2 + B^2/\mu_0\}$.

$P^\alpha = 1/c^2 \int S^\alpha d^3x, \quad P^4 = 1/c \int w\, d^3x = E/c$.

3.1.6.7. Energie-Impuls-Tensor in Medien. Für freie Felder ohne Ströme erhält man aus $L = -\int H^{ik} dB_{ik}/2$ mit $H^{ik} = \lambda^{[ik][rs]} B_{rs}$ und $F^{i[jk]} = -A^i H^{jk}$ den (unsymmetrischen) *Minkowski-Tensor* $T_i^{(M)k} = -(1/4)\,\delta_i^k H^{rs} B_{rs} + H^{rk} B_{ri}$.
Symmetrisch ist der *Abraham-Tensor*

$T_i^{(A)k} = 1/2\,(B_{ir} H^{kr} + B^{kr} H_{ir}) - 1/4\,\delta_i^k B_{rs} H^{rs} + u^l/(2c^2)$
$\{u_i [B^{kr} H_{lr} - H^{kr} B_{lr}] + u^k [B_{ir} H_l^r - H_{ir} B_l^r]\}$.

Streng symmetrisch muß nur die Summe von Feld und Materietensor sein. Die Aufspaltung ist weitgehend willkürlich. Hauptargumente für den Minkowski-Tensor sind das Auftreten von Drehmomenten in Medien (magn. Kristallkugel im homogenen Magnetfeld) und das Transformationsverhalten einer Welle, die sich mit $c^* = c/n$ ausbreitet, verglichen mit dem eines wechselwirkungsfreien Teilchens der gleichen Geschwindigkeit.

3.1.7. Kontinuumsmechanik und Thermodynamik

3.1.7.1. Begriffsbildung. In der Anfangszeit der Speziellen Relativitätstheorie faßte man thermodynamische Größen, wie Temperatur, innere Energie, als zu transformierende Gebilde auf: $T = T_0\sqrt{1-(v/c)^2}$, $E = E_0/\sqrt{1-(v/c)^2}$.
Günstiger für die Kontinuumsmechanik ist die Einführung dieser Größen als Invarianten, die im lokalen momentanen Ruhsystem definiert sind, dort ihre physikalische Bedeutung haben und gemessen werden (ähnlich Eigenzeit und Ruhmasse). Für eine einkomponentige Flüssigkeit als Beispiel haben wir in jedem Punkt die Temperatur $T(x)$, den Druck $p(x)$, die Teilchenzahldichte $n(x)$ (Mol/cm³), das spezifische Volumen $v = 1/n$, die molare Entropie $S(x)$ und als Zustandsgleichung die molare innere Energie $E(S, n)$ als Potential mit $T = \partial E/\partial S$, $p = n^2\,\partial E/\partial n$. Die Verhältnisse sind so, daß jede Mediumszelle in guter Näherung als kleines homogenes System angesehen werden kann. Als kinematische Größe tritt noch $u^i(x)$ hinzu.
Wegen der Bedeutung des lokalen Ruhsystems benötigt man ein Verfahren, einen Vektor so zu zerlegen, daß seine Teile gerade die zeitliche und räumliche Komponente in diesem System sind. Das leistet der „räumliche Projektionstensor" $h_k^i = \eta_k^i + u^i u_k/c^2$, $V^i = h_k^i V^k - u^i (u_k V^k)/c^2$, was eine Zerlegung in Richtung von u^i und senkrecht dazu bewirkt.

3.1.7.2. Teilchenzahlerhaltung. Mittels Gaußschem Satz beweist man, daß $0 \equiv (nu^i)_{,i} = (nu^\alpha)_{,\alpha} + (nu^4)_{,4} \equiv (\tilde{n}v^\alpha)_{,\alpha} + \partial\tilde{n}/\partial t$ Teilchenzahlerhaltung bedeutet. $\tilde{n} = n\,dt/d\tau = n/\sqrt{1-(v/c)^2}$. Die Teilchenzahl ist $N = \int dN = \int \tilde{n}\,d^3x = \int n\,d^{(3)}x$. $d^{(3)}x$ ist das räumliche Volumenelement des lokalen Ruhsystems. Ausgeschlossen werden dadurch chemische Umsetzungen, Zerstrahlung usw.

3.1.7.3. Bewegungsgleichung und 1. Hauptsatz. Für jedes kontinuumsmechanisch-thermodynamisches System existiert ein divergenzfreier Energie-Impuls-Tensor, zusammengesetzt aus einem mechanischen und einem elektromagnetischen Anteil $T^{ik} = T^{ik}_{\text{mech}} + T^{ik}_{\text{elm}} = T^{(ik)}$, $T^{ik}_{,k} = 0$. Zerlegung bezüglich u^i gibt:

$T^{ik}_{\text{mech}} = 1/c^2\,\{nEu^i u^k + I^i u^k + \bar{I}^k u^i\} - \sigma^{ik}$.
$I^i u_i = \bar{I}^i u_i = \sigma^{ik} u_k = \sigma^{ik} u_i = 0$.

$\mu = nE/c^2$ ist die Massendichte, I^i die Impuls- und $\bar{I}^i$ die Energiestromdichte. Für Flüssigkeiten ist der Spannungstensor $\sigma_{ik} = -ph_{ik} + \sigma^{\text{Reibg.}}_{ik} = -ph_{ik} + 2\eta_{\text{I}} h_i^r h_k^s u_{(r,s)} + \eta_{\text{II}} h_{ik} u^r_{,r}$. $T^{ik}_{,k} = 0$ liefert $T^{ik}_{\text{mech},k} = -T^{ik}_{\text{elm},k} \equiv f^i$. f^i ist die bekannte Kraftdichte. Überschrieben mit h_l^a bzw. u_i liefern *mechanische Bewegungsgleichung* und *1. Hauptsatz.*

$\mu\dot{u}^i = h^i_j\{f^j + \sigma^{jk}_{,k} - (\dot{I}^j + I^j u^k_{,k} + u^j_{,k}\bar{I}^k)/c^2\}$ (Bewegungsgleichung).

$(nEu^k)_{,k} = n\dot{E} = -\bar{I}^k_{,k} - u_i(\sigma^{ik}_{,k} + f^i - \dot{I}^i/c^2)$ (1. Hauptsatz).
Die Änderung der inneren Energie setzt sich aus der je Zeiteinheit zugefügten Wärmemenge (Energiefluß durch Berandung) und der Leistung der mechanischen und elektromagnetischen Kräfte zusammen. Dazu wird auch das letzte Glied gezählt, das aus der Trägheit der Energie resultiert.

3.1.7.4. Entropiesatz. Die Entropie im üblichen Sinne hängt mit der 4. Komponente des *Entropiestromvektors zusammen,*

$S^i = \hat{S}^i + nSu^i/c$, $\hat{S}^i = h^i_k S^k = I^i/T$.

S bzw. nS sind Entropie je Mol bzw. Volumeneinheit im lokalen Ruhsystem. $\hat{S}^i$ ist die Dreierentropiestromdichte. Sie hängt mit der auf der Temperatur T zugeführten Wärme (Energie) zusammen. Der *2. Hauptsatz* lautet damit $S^i_{,i} = \eta \geq 0$. η ist die Entropieproduktionsdichte. Der 2. Hauptsatz mit $\eta \geq 0$ besagt, daß Entropie erzeugt aber nicht vernichtet werden kann. Sie kann höchstens wegströmen.
Für ein System ohne elektromagnetisches Feld sollen die Ausdrücke des 2. Hauptsatzes im einzelnen berechnet werden. Wegen $I^i = \bar{I}^i$ entsteht mit $\dot{E} = T\dot{S} - p\dot{v}$, $\sigma^{ik} = -ph^{ik} + \sigma^{ik}_{\text{Reibg.}}$ und Einsetzen des 1. Hauptsatzes:

$$S^i_{,i} \equiv (I^i/T + nSu^i)_{,i} = \eta$$
$$= 1/T\left\{\sigma^{ik}_{\text{Reibg.}}\, u_{(i,k)} - \left(\frac{T_{,i}}{T} + \frac{\dot{u}_i}{c^2}\right) I^i\right\} = \sum_A J_A X_A \geq 0.$$

η wird nach den Ideen ONSAGERS als $\sum$ Ströme × Kräfte interpretiert, wofür man in Gleichgewichtsnähe die linearen Ansätze erhält. $J_A = \sum_B L_{AB} X_B$. Entsprechend der Curieschen Regel entsteht im einzelnen:

$I_i = -\varkappa h_i^k (T_{,k} + T\dot{u}_k/c^2)$ relativistische Wärmeleitung,

$\sigma^{ik}_{\text{Reibg.}} = \lambda\{h^{ir} h^{ks} u_{(r,s)} - (1/3)\,h^{ik} u^r_{,r}\}$ relativistische innere Reibung.

$\eta \geq 0$ verlangt $\varkappa, \lambda \geq 0$.

3.2. Allgemeine Relativitätstheorie

3.2.1. Physikalische Einführung

Während die Spezielle Relativitätstheorie (SRT) ihre Entstehung A. EINSTEIN, H. A. LORENTZ, H. POINCARÉ und H. MINKOWSKI verdankt, ist die Schaffung der Allgemeinen Relativitätstheorie (ART) im wesentlichen das Werk EINSTEINS selbst, der etwa acht Jahre lang hart um die Entwicklung dieser Theorie – oft in sehr widersprüchlichen Gedan-

kengängen – gerungen hat. Die endgültige Fassung gelang ihm schließlich 1915. Sie wurde in den Annalen der Physik, Band 41 (1916), S. 769–822, unter dem Titel „Die Grundlagen der allgemeinen Relativitätstheorie" publiziert.
Die SRT hat sich als eine tiefgründige Theorie mit prinzipiell neuen Einsichten in das Wesen von Raum und Zeit ausgezeichnet bewährt. Die Punktmechanik, Kontinuumsmechanik (Elastomechanik und Hydromechanik), Elektromagnetik, Optik, Quantenmechanik und Quantenfeldtheorie (zum Teil auch die Thermodynamik und Statistik) konnten auf ihrer Basis neu fundiert werden, was zur Erklärung und zum Verständnis einer kaum überschaubaren Zahl physikalischer Effekte führte.
Trotz dieser großartigen Erfolge der SRT erkannte EINSTEIN bald prinzipielle Einschränkungen und Grenzen dieser Theorie, die für den weiteren Fortschritt der Physik überwunden werden mußten oder sogar noch bis heute als Barrieren vor uns stehen. Zunächst fielen zweierlei tiefgreifende Mängel auf: Erstens sind in der SRT die Physiker als erkennende und messende Forscher („Beobachter") bei der Beschreibung von Naturphänomenen auf die Plazierung in Inertialsystemen festgelegt. In unserer wirklichen Welt gibt es aber in Strenge überhaupt keine Inertialsysteme, die ja eine Idealisierung allgemeiner Bezugssysteme mit Beschleunigung und Rotation (Nichtinertialsysteme) sind.
Die typischen Bezugssysteme in unserer realen Welt sind also die Nichtinertialsysteme. Da aber solche im Rahmen der SRT nicht zulässig sind, ergab sich für Einstein die dringende Forderung, die SRT so zu verallgemeinern, daß sie für Beobachter in Nichtinertialsystemen anwendbar ist. Dieses Programm wurde dann in der ART realisiert.
Zweitens ist der mathematische Apparat der SRT, deren geometrische Grundlage der vierdimensionale ungekrümmte Minkowski-Raum (ebene Raum-Zeit) ist, auf die Benutzung der Galilei-Koordinaten x, y, z, ct (vierdimensionale Verallgemeinerung der kartesischen Koordinaten) zugeschnitten. Die Lösung praktischer Probleme ist aber in der Regel an solche Koordinaten gebunden, die der Symmetrie des behandelten Objekts optimal angepaßt sind. Da es sich dabei i. allg. um die Forderung nach Verwendbarkeit krummliniger Koordinaten handelt, war also auch unter diesen Aspekten die Erweiterung der SRT zu fordern. Dieser Gesichtspunkt wurde ebenfalls in der ART realisiert.
Bei der obigen Aufzählung der in die SRT einbezogenen Gebiete wurde ein fundamentaler Gegenstand, nämlich die Gravitationstheorie, nicht erwähnt. Die Ursachen dafür liegen im folgenden:
Die physikalische Erkenntnis des Naturphänomens der Gravitation (Schwerkraft) verdankt die Menschheit bekanntlich ISAAC NEWTON, der das Attraktionsgesetz für die Schwerkraft zwischen Massen entdeckt hat. NEWTONS Einsicht, daß es sich bei der Schwerkraft um ein universelles Phänomen handelt, dem sowohl irdische Massen als auch kosmische Massen unterliegen, war epochal. Insbesondere war erst dadurch eine quantitative Aussage über die Planetenbewegung möglich geworden.
Es ist sofort verständlich, daß nach Schaffung der SRT EINSTEINS großes Interesse darauf gerichtet war, die nichtrelativistische Newtonsche Gravitationstheorie so zu verallgemeinern, daß sie dem Speziellen Relativitätsprinzip genügen, also in den Rahmen der SRT passen würde. Mit dieser Aufgabe waren damals mehrere Forscher beschäftigt. Dabei gelang es zwar, eine speziell-relativistische Verallgemeinerung der Newtonschen Gravitationstheorie zu finden, aber die Voraussagen dieser Theorie, z. B. bei der Periheldrehung des Merkur (langsame Drehung der Kepler-Ellipsen), standen im Gegensatz zur Erfahrung. Diese widersprüchliche Situation ließ in EINSTEIN nach und nach die Überzeugung reifen, daß der Rahmen der SRT für die Erfassung des Phänomens der Gravitation zu eng ist. Vielmehr schien es so, daß die Gravitation tief in die Geometrie von Raum und Zeit eingreift. Durch diese Erkenntnis war ein weiterer Mangel der SRT aufgedeckt.
Über diese theoretisch-physikalischen Aspekte hinausgehend, drängten sich EINSTEIN, dessen philosophisches Weltverständnis stark durch B. SPINOZA und E. MACH geprägt war, auch prinzipielle philosophische Mängel der SRT auf, die man kurz folgendermaßen skizzieren kann: Die SRT führte zur Liquidierung des insbesondere von MACH kritisierten Newtonschen Fundaments eines absoluten Raumes und einer absoluten Zeit und schuf die vierdimensionale Raum-Zeit als Arena des physikalischen Geschehens, wodurch Raum und Zeit für sich zu relativen Kategorien erklärt wurden. Dieser erkenntnistheoretische Schritt ist sicher als ein echter Fortschritt zu werten. Aber diese Position konnte EINSTEIN noch nicht befriedigen. Jetzt war nämlich diesem neuen Begriff der Raum-Zeit eine ähnliche philosophische Absolutheit zuzuschreiben wie Raum und Zeit einzeln bei NEWTON. Es konnten damit gegen die neu entdeckte Raum-Zeit dieselben Bedenken vorgebracht werden wie früher gegen Raum und Zeit für sich. In erster Linie war dabei an die Machsche These zu denken, daß die Begriffe von Raum und Zeit ohne Bezug auf die in Raum und Zeit bewegten Körper inhaltsleer seien. Vielmehr sollte nach MACH die Existenz von Raum und Zeit aus der Existenz der Körper (historisch bedingte Simplifizierung unseres heutigen Materiebegriffes) resultieren.
Diese Vermutung MACHS ist eine wichtige Komponente des sogenannten Machschen Prinzipes, dessen tiefer Sinn bis heute noch umstritten ist. Offensichtlich handelt es sich bei diesen Ideen um ein interessantes dialektisches Gedankengut, das EINSTEINS Denken nachdrücklich geformt und ihn stark bei seinem Suchen nach der ART inspiriert hat.

3.2.2. Allgemeines Relativitätsprinzip

Abgesehen von der Problematik um die Quantentheorie und Elementarteilchenphysik, worauf hier nicht näher eingegangen werden kann, sind die Grenzen der SRT entsprechend den obigen Ausführungen durch die folgenden Barrieren zu charakterisieren: Festlegung auf Inertialsysteme, Beschränkung des mathematischen Apparates auf Galilei-Koordinaten, Nichterfassung der Gravitation.
Die ersten beiden Problemkreise sind ganz eng miteinander verbunden, da – in vierdimensionaler Sicht – die Einbeziehung nichtinertialer Bewegungszustände von Bezugssystemen und die Benutzung krummliniger Koordinaten auf derselben Ebene mathematischer Behandlung liegen. Dieser Gegenstand soll im folgenden behandelt werden. In dem dabei gesteckten mathematischen und physikalischen Rahmen wird dann die Abhandlung der Einsteinschen Gravitationstheorie ermöglicht, die wegen ihrer Wichtung später gesondert dargestellt werden wird.
Ein Leitstern bei der Entwicklung der ART war für EINSTEIN das Allgemeine Relativitätsprinzip, das als eine inhaltliche Weiterentwicklung des Speziellen Relativitätsprinzipes aufgefaßt wird. Wie bereits ausgeführt, trifft das Spezielle Relativitätsprinzip die Aussage, daß die im Rah-

men der SRT erfaßbaren physikalischen Naturgesetze für alle relativ zueinander (geradlinig und gleichförmig) bewegten Inertialsysteme bei Benutzung von Galilei-Koordinaten dieselbe Form haben. Diese tiefgründige Aussage über die Form von Naturgesetzen legt die Frage einer eventuellen Verallgemeinerung auf Nichtinertialsysteme nahe. Bei der Behandlung dieser Frage muß man sich zweckmäßigerweise noch einmal die volle Tragweite des Speziellen Relativitätsprinzips vor Augen führen:

Dieses Prinzip impliziert, daß in die physikalischen Grundgesetze keine frei wählbare Geschwindigkeit eingehen darf, die man sonst als absolute Geschwindigkeit des betrachteten Inertialsystems gegenüber einem naturgegebenen Bezug (z. B. dem früher angenommenen, aber durch die SRT widerlegten Äther) zu interpretieren hätte. Dieser Tatbestand besagt, daß der Ablauf der physikalischen Vorgänge in relativ zueinander bewegten Inertialsystemen bei Wahl derselben Anfangs- und Randbedingungen in gleicher Weise vonstatten geht. Man sagt deshalb, daß alle Inertialsysteme in diesem Sinne physikalisch äquivalent sind, so daß keines von ihnen vor den anderen ausgezeichnet ist. Die Inertialsysteme haben also eine doppelte, koinzidierende Äquivalenz:

a) Äquivalenz bezüglich der Forminvarianz der Naturgesetze. Für diese Eigenschaft wird hier der Begriff „Kovarianz-Äquivalenz“ benutzt, da die Begriffe Kovarianz und Forminvarianz synonym sind.
b) Äquivalenz bezüglich des bildlichen Ablaufes des physikalischen Geschehens, die mit „Prozeß-Äquivalenz“ bezeichnet wird. Da es in der ART um beliebige vierdimensionale Koordinaten-Transformationen geht, denn nur so läßt sich die Einschränkung auf Inertialsysteme mit Galilei-Koordinaten überwinden, gelangt man also unmittelbar zur Frage der Forminvarianz der physikalischen Grundgesetze bei beliebigen (kontinuierlichen) vierdimensionalen Koordinaten-Transformationen. In direkter Verallgemeinerung der Idee des Speziellen Relativitätsprinzips wird man damit zur nachfolgenden Formulierung geführt.

Allgemeines Relativitätsprinzip:
„Die Grundgesetze der Physik besitzen für zwei in beliebigem Bewegungszustand befindliche Beobachter bei Benutzung beliebiger, kontinuierlich auseinander hervorgehender Koordinatensysteme dieselbe Form“.

Da in diesem Sinne das Allgemeine Relativitätsprinzip eine Aussage über die Forminvarianz (Kovarianz) der Grundgesetze der Physik trifft, spricht man oft auch synonym vom Allgemeinen Kovarianzprinzip. Die in diesem Prinzip stekkende Gleichberechtigung der Koordinatensysteme ist mithin eine Gleichberechtigung hinsichtlich der Form physikalischer Naturgesetze. Da beliebige vierdimensionale Koordinaten-Transformationen den Übergang zu beliebig bewegten Bezugssystemen induzieren, resultiert daraus eine Gleichberechtigung hinsichtlich der Form für beliebige Bezugssysteme. Das Allgemeine Relativitätsprinzip genügt also der oben erklärten Kovarianz-Äquivalenz.

Wesentlich anders ist aber die tägliche Erfahrung hinsichtlich der ebenfalls oben eingeführten Prozeß-Äquivalenz. Von einem rotierenden Karussell aus gesehen, erscheint der prozessuale Ablauf eines physikalischen Geschehens bildlich ganz anders als von einem Inertialsystem aus. In allgemein-relativistischen Bezugssystemen besteht demnach keineswegs Prozeß-Äquivalenz des Geschehens. Im Unterschied zur SRT fallen also beide Äquivalenzen in der ART klar auseinander. Die Frage nach der Ursache für diesen Tatbestand mündet in den Problemkreis um das Machsche Prinzip ein.

Die Wertung des Allgemeinen Relativitätsprinzipes räumt diesem Prinzip einen fundamentalen Erkenntniswert ein, da es Grundsatzaussagen über die Form der physikalischen Naturgesetze macht. Es ist als Kernstück der ART anzusehen, die nicht mit der Einsteinschen Gravitationstheorie identifiziert werden soll, denn letztere erscheint im Rahmen der ART als eine Theorie – mit allerdings besonderer Wichtung – neben den anderen allgemein-relativistischen Theorien (Mechanik, Elektromagnetik, Quantentheorie usw.). Natürlich bekommt die ART erst durch die Einsteinsche Gravitationstheorie mit deren Erkenntnis von der Krümmung der Raum-Zeit ihr spezifisches Zentrum. Inhalt und Deutung des Allgemeinen Relativitätsprinzips waren vom Anfang an umstritten. Die Diskussion darum dauert bis heute an.

Um der geforderten Kovarianz bei der Formulierung der physikalischen Grundgesetze gerecht zu werden, benötigt man einen spezifischen mathematischen Kalkül (z. B. für physikalische Tensorfelder den Tensorkalkül). Dieser Kalkül wurde für die Geometrie gekrümmter Räume im Anschluß an C. F. GAUSS, N. I. LOBATSCHEWSKY und J. BOLYAI von B. RIEMANN, E. CHRISTOFFEL, C. RICCI und T. LEVI-CIVITA entwickelt. EINSTEIN erhielt bei der Nutzbarmachung dieses sogenannten Ricci-Kalküls für die Physik eine große Hilfe durch M. GROSSMANN.

3.2.3. Äquivalenzprinzip von schwerer und träger Masse

Das Newtonsche Bewegungsgesetz (Lex secunda) hat für die Entwicklung der ART eine große historische Bedeutung. Es soll deshalb hier noch einmal darauf eingegangen werden. Seine Formulierung lautet: Wirkt auf einen betrachteten Massenpunkt mit der trägen Masse m_t und der Geschwindigkeit $\boldsymbol{v}(t)$, wobei t die absolute Zeit nach NEWTON ist, eine eingeprägte Kraft $\boldsymbol{F}(t)$, so bewegt sich in einem Inertialsystem der Massenpunkt gemäß dem Bewegungsgesetz $m_t \, \mathrm{d}\boldsymbol{v}(t)/\mathrm{d}t = \boldsymbol{F}(t)$. Handelt es sich bei der eingeprägten Kraft $\boldsymbol{F}$ speziell um die Newtonsche Gravitationskraft, die (ihrem Betrage nach) zwischen zwei schweren Massen m_s und M_s vom Abstand r durch $G = \gamma_N (m_s M_s)/4\pi r^2$ (γ_N = Newtonsche Gravitationskonstante) gegeben ist, so resultiert aus den beiden Formeln die Beziehung $m_t \, \mathrm{d}v/\mathrm{d}t = \gamma_N \, r(m_s M_s/4\pi r^3)$, wenn dem betrachteten Massenpunkt die schwere Masse m_s zugeordnet wird, der von einem anderen Massenpunkt mit der schweren Masse M_s gravitativ angezogen werden möge. Wird auf diese Beziehung die von GALILEI bei der Formulierung seiner Fallgesetze besonders herausgestellte, aber vor ihm bereits bekannte empirische Aussage angewendet, daß die Beschleunigung eines fallenden Körpers von der Masse des fallenden Körpers unabhängig ist, so heißt das m_s/m_t = Konstante. Ohne Beschränkung der Allgemeinheit lassen sich nun die Maßeinheiten der trägen und schweren Masse so festlegen, daß $m_s = m_t$ gesetzt werden kann.

Diese Gleichung drückt die fundamentale Erkenntnis der Gleichheit (Äquivalenz) von träger und schwerer Masse aus. Träge und schwere Masse erscheinen damit wesensgleich. Während diese Gleichheit von R. v. EÖTVÖS und P. ZEEMAN in den Jahren 1917 bis 1922 mit einer relativen Genauigkeit von $5 \cdot 10^{-8}$ experimentell bestätigt werden konnte, weisen Experimente aus neuerer Zeit eine relative Genauigkeit von 10^{-12} aus.

Da der Begriff der trägen Masse auf die Kinematik hinweist, der Begriff der schweren Masse aber der Gravitation zugeordnet ist, wird die Gleichheit von träger und schwerer Masse in der lokalen (kleine Raumgebiete und kurze Zeitintervalle) Ununterscheidbarkeit von kinematischer Beschleunigung und Gravitationsbeschleunigung widergespiegelt (Prinzip von der lokalen Äquivalenz von beschleunigtem Bezugssystem und korrespondierendem Gravitationssystem). Zur Verdeutlichung denke man z. B. an die Gleichartigkeit des physikalischen Geschehens in einem beschleunigten Fahrstuhl und in einem geeignet gewählten homogenen Gravitationsfeld.

Die empirisch konstatierte Gleichheit von träger und schwerer Masse war für EINSTEIN ein entscheidender Fingerzeig bei der Entwicklung seiner Gravitationstheorie – im besonderen bei der Aufstellung seiner Bewegungsgleichung für einen Probekörper. Er konnte diese Forderung in ganz natürlicher Weise dadurch erfüllen, daß er als Bewegungsgleichung die Gleichung einer Geodäten (kürzeste Linie) in der gekrümmten Raum-Zeit wählte. Auf diese Weise kam derselbe Massenfaktor bei der kinematischen und gravitativen Beschleunigung, die sich – streng genommen – nicht mehr trennen lassen, zu stehen.

3.3. Gravitationstheorie

3.3.1. Newtonsche Gravitationstheorie

NEWTON erkannte das grundlegende Gesetz der Gravitation: Zwei *Punktmassen* m_1 und m_2 ziehen sich gegenseitig mit einer Kraft $\boldsymbol{F} = \gamma m_1 m_2 \boldsymbol{r}/r^3$ an, die längs der Verbindungsgeraden wirkt und betragsmäßig dem Quadrat des Abstandes umgekehrt proportional ist. Die *Newtonsche Gravitationskonstante* γ ergibt sich zu $\gamma = 6{,}67 \cdot 10^{-8}\,\mathrm{cm^3\,g^{-1}\,sec^{-2}}$.

Wenn eine Kraft in Abhängigkeit vom Ort gegeben ist, spricht man von einem Kraftfeld, in unserem Fall vom *Gravitations-* oder *Schwerefeld.*

Mit der Newtonschen Lex secunda $m\,\mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2 = \boldsymbol{F}$ und dem Kraftgesetz läßt sich die Bewegung von Massen im Schwerefeld der Erde oder der Sonne recht genau beschreiben. Die Keplerschen Gesetze der Planetenbewegung können aus den beiden o. g. Gleichungen abgeleitet werden.

Massen bewegen sich unter dem Einfluß der Schwerkraft, die ihrerseits durch Massen hervorgerufen wird. Eine Punktmasse, deren Beitrag zum Gravitationsfeld vernachlässigt werden kann, bezeichnet man als *Testteilchen*; es bewegt sich im Feld, das von den anderen Massen erzeugt wird und dient als Sonde zur Ausmessung dieses Feldes.

Der Übergang von felderzeugenden Punktmassen zu *kontinuierlich* verteilter Masse führt zu folgender Formulierung der Newtonschen Gravitationstheorie:

$$\Delta U = 4\pi\gamma\mu,$$
$$\mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2 = -\operatorname{grad} U.$$

Dabei bezeichnet Δ den Laplace-Operator des dreidimensionalen Euklidischen Raumes. Die Massendichte μ und das *Newtonsche Gravitationspotential* U hängen i. allg. von Ort und Zeit ab, $U = U(\boldsymbol{r}, t)$, $\mu = (\boldsymbol{r}, t)$. Bei vorgegebener Massendichte ist das Gravitationspotential U aus der partiellen Differentialgleichung $\Delta U = 4\pi\gamma\mu$ (Poisson-Gleichung) zu bestimmen. Der negative Gradient von U ergibt die Schwerkraft bezogen auf die Masse eines Testteilchens, dessen Bewegung durch die Gleichung $\mathrm{d}^2\boldsymbol{r}/\mathrm{d}t^2 = -\operatorname{grad} U$ beschrieben wird.

Bei der Behandlung zahlreicher Probleme hat sich die Newtonsche Gravitationstheorie glänzend bewährt. Ihr entscheidender Mangel besteht jedoch darin, daß es sich um eine *Fernwirkungstheorie* handelt. Wenn sich die Massenverteilung in einem Raumgebiet ändert, so wirkt sich dies *momentan* im Schwerefeld an beliebig weit davon entfernten Punkten des Raumes aus. Im Gegensatz zur Speziellen Relativitätstheorie existiert also keine endliche Grenzgeschwindigkeit. (Mathematisch drückt sich das darin aus, daß der Differentialoperator in $\Delta U = 4\pi\gamma\mu$ keine Zeitableitung enthält.)

Es erwies sich als unmöglich, eine speziell-relativistische Gravitationstheorie aufzustellen, die in sich widerspruchsfrei ist und die Beobachtungsdaten richtig wiedergibt (vgl. 3.2.1.).

3.3.2. Einsteinsche Gravitationstheorie

Mit der Allgemeinen Relativitätstheorie (vgl. 3.2.) schuf ALBERT EINSTEIN zugleich eine neue Gravitationstheorie. Sie beruht auf einer anderen begrifflichen Grundlage als die Newtonsche Theorie, enthält diese aber als Spezialfall.

In einem frei fallenden Fahrstuhl oder in einem antriebslos fliegenden Raumschiff kompensieren sich Gravitations- und Beschleunigungseffekte *lokal*, d. h. bei Beschränkung auf hinreichend kleine räumliche und zeitliche Abstände. Auch in einem Gravitationsfeld sind lokal die Gesetze der Speziellen Relativitätstheorie gültig; man kann ein *lokales Inertialsystem* einführen. Wenn sich die Beobachtungen jedoch über größere Abstände erstrecken, werden die Abweichungen von der speziell-relativistischen Physik immer stärker. Das lokale Inertialsystem gestattet keine adäquate Beschreibung der Gravitation. Beschleunigte Bewegung ist im Rahmen der Speziellen Relativitätstheorie erfaßbar, Gravitation nicht. Um dieses Phänomen zu beschreiben, ist nach EINSTEIN der Übergang vom Minkowski-Raum zu einer *gekrümmten* Raum-Zeit zu vollziehen. Dieser vierdimensionale *Riemannsche Raum* ist mit einer *Metrik* versehen. Durch die zehn Komponenten g_{ij} des metrischen Tensors ist das Quadrat des Abstandes zwischen infinitesimal benachbarten Punkten (mit den Koordinaten x^i und $x^i + \mathrm{d}x^i$) festgelegt: $\mathrm{d}s^2 = g_{ij}\,\mathrm{d}x^i\,\mathrm{d}x^j$.

Gemäß der Einsteinschen Summenkonvention ist hier und in den folgenden Formeln über doppelt auftretende Indizes von 1 bis 4 zu summieren.

Um die Krümmung eines Riemannschen Raumes zu erläutern, geht man von einer Fläche im Euklidischen Raum aus. Man wählt auf dieser Fläche einen Punkt P und betrachtet die Schnittkurven der Fläche mit allen Ebenen, die den Flächennormalenvektor $\boldsymbol{n}$ enthalten. Im allgemeinen existieren zwei derartige Kurven, für welche die Krümmungsradien R_1 und R_2 extremal sind. Die *Gaußsche Krümmung* K einer Fläche in einem Flächenpunkt ist definiert als der Kehrwert des Produkts der extremalen Krümmungsradien, $K = 1/(R_1 R_2)$. Die (innere) Geometrie einer gekrümmten Fläche ($K \neq 0$) ist nichteuklidisch.

Der *Krümmungstensor* R^k_{lmn} eines Riemannschen Raumes stellt die Verallgemeinerung des Begriffes der Gaußschen Krümmung auf höhere Dimensionszahl dar.

Die Gravitation äußert sich in einer Krümmung der Raum-Zeit ($R^k_{lmn} \neq 0$), die Gravitation ist geometrisiert. Wenn Gravitationseffekte vernachlässigbar sind, darf man die (ungekrümmte) Raum-Zeit der Speziellen Relativitätstheorie

zugrunde legen. Die Bewegung von Testteilchen im Gravitationsfeld erfolgt auf *Geodäten.* (In der Riemannschen Geometrie sind Geodäten die kürzesten Verbindungslinien zwischen zwei Punkten.) Anstelle von $\mathrm{d}^2 \mathbf{r}/\mathrm{d}t^2 = -\mathrm{grad}\, U$ in der Newtonschen Theorie tritt also jetzt die Geodätengleichung

$$\frac{\mathrm{d}^2 x^i}{\mathrm{d}\tau^2} + \left\{ {i \atop jk} \right\} \frac{\mathrm{d}x^j}{\mathrm{d}\tau} \frac{\mathrm{d}x^k}{\mathrm{d}\tau} = 0 .$$

Dabei bedeutet τ die *Eigenzeit* des frei fallenden Teilchens, und die Christoffel-Symbole $\left\{ {i \atop jk} \right\}$ sind aus dem metrischen Tensor und seinen partiellen Ableitungen gebildet:

$$\left\{ {i \atop jk} \right\} = \frac{1}{2} g^{im} \left(\frac{\partial g_{mk}}{\partial x^j} + \frac{\partial g_{mj}}{\partial x^k} - \frac{\partial g_{jk}}{\partial x^m} \right).$$

Die Lösungen $x^i = x^i(\tau)$ der Geodätengleichung ergeben die *Weltlinien* des Teilchens unter bestimmten Anfangsbedingungen für Ort und Geschwindigkeit. Das Studium der Geodäten gibt Aufschluß über die Struktur der Raum-Zeit. Insbesondere bewirkt die Krümmung eine Abweichung zwischen benachbarten Geodäten. Das Gesetz $\Delta U = 4\pi\gamma\mu$, wonach die Massen das Gravitationspotential in der Newtonschen Theorie bestimmen, wird durch die *Einsteinschen Feldgleichungen* $R_{ij} - (1/2)\, R g_{ij} = \varkappa T_{ij}$ ersetzt. Die Größen R_{ij} und R sind Bildungen aus Krümmungstensor und metrischem Tensor, so daß in die linke Seite der Feldgleichungen nur geometrische Größen eingehen, nämlich die zehn Komponenten g_{ij} und ihre partiellen Ableitungen bis zur 2. Ordnung.

Der (symmetrische) *Energie-Impuls-Tensor* auf der rechten Seite der Einsteinschen Feldgleichungen stellt den Quellterm des Gravitationsfeldes dar. Seinen zehn Komponenten T_{ij} sind Massen- bzw. Energiedichte, Impuls- und Energiestromdichte sowie Spannungen eines mechanischen Kontinuums oder eines beliebigen Feldes (außer dem Gravitationsfeld selbst) zugeordnet. Die Einsteinsche Energie-Masse-Relation (vgl. 3.1.4.) schließt prinzipiell die Beschränkung auf rein mechanische Quellterme aus. Auch das elektromagnetische Feld bewirkt eine Krümmung der Raum-Zeit und damit ein Gravitationsfeld.

In den Raum-Zeit-Gebieten außerhalb von Quellen, d. h. für $T_{ij} = 0$, gehen die Einsteinschen Feldgleichungen in die *Vakuumfeldgleichungen* $R_{ij} = 0$ über.

Im Newtonschen Grenzfall können gegenüber der Massendichte alle übrigen Bestandteile von T_{ij} vernachlässigt werden, und die zehn Einsteinschen Feldgleichungen reduzieren sich auf eine einzige Gleichung. Aus dem Vergleich mit $\Delta U = 4\pi\gamma\mu$ ergibt sich für die *Einsteinsche Gravitationskonstante* $\varkappa = 8\pi\gamma/c^4$.

Die Bewegungsgleichung $\frac{\mathrm{d}^2 x^i}{\mathrm{d}\tau^2} + \left\{ {i \atop jk} \right\} \frac{\mathrm{d}x^j}{\mathrm{d}\tau} \frac{\mathrm{d}x^k}{\mathrm{d}\tau} = 0$ eines Testteilchens läßt sich aus den nichtlinearen Feldgleichungen $R_{ij} - (1/2)\, R g_{ij} = \varkappa T_{ij}$ herleiten und braucht deshalb nicht zusätzlich postuliert zu werden.

Bei vorgegebener Struktur von T_{ij} bilden die Einsteinschen Feldgleichungen ein nichtlineares partielles Differentialgleichungssystem für g_{ij} und die Materiefelder in T_{ij}.

Strenge Lösungen wurden meist unter speziellen Symmetrieannahmen gewonnen. Das einzige *kugelsymmetrische Vakuumfeld* ($T_{ij} = 0$) ist die (äußere) *Schwarzschild-Lösung*

$$\mathrm{d}s^2 = (1 - 2m/r)^{-1} \mathrm{d}r^2 + r^2 (\mathrm{d}\vartheta^2 + \sin^2\vartheta\, \mathrm{d}\varphi^2) - (1 - 2m/r)\, \mathrm{d}t^2 .$$

Dabei bezeichnen ϑ und φ Winkelkoordinaten auf der Kugeloberfläche, r eine radiale und t eine zeitliche Koordinate (für $r > 2m$). Das von der Sonne erzeugte Gravitationsfeld wird außerhalb der Sonne durch die Schwarzschild-Lösung beschrieben. Aus der Geodätengleichung in dieser Raum-Zeit gewinnt man Aussagen über die *Periheldrehung* der Planeten oder die *Lichtablenkung* am Sonnenrand. Diese klassischen Effekte wurden von der Einsteinschen Theorie vorausgesagt und dann bestätigt (vgl. 3.4.).

Um das Gravitationsfeld im Inneren und Äußeren eines kugelsymmetrischen Himmelskörpers (z. B. eines Neutronensternes) zu beschreiben, ist bei einem Radius r_0 ($r_0 > 2m$) eine *innere Lösung* (mit $T_{ij} \neq 0$) anzuschließen, wodurch die *Singularität* bei $r = 0$ vermieden wird.

Die Schwarzschild-Lösung weist außer dieser Singularität eine weitere Besonderheit auf: Wenn ein Testteilchen in das Raum-Zeit-Gebiet mit $r < 2m$ gelangt ist, kann es daraus nie mehr entweichen; es bleibt in dem starken Gravitationsfeld eingefangen. Deshalb spricht man von einem *Schwarzen Loch,* und die Fläche $r = 2m$ heißt *Horizont.* Unter extremen Bedingungen läßt sich der Gravitationskollaps eines Sternes nicht aufhalten, die Sternoberfläche verschwindet unter dem Horizont; es bildet sich ein Schwarzes Loch. Ob diese theoretisch möglichen Endzustände der Sternentwicklung tatsächlich existieren, ist noch nicht endgültig geklärt.

Strenge Lösungen der Einsteinschen Feldgleichungen gibt es für *stationäre axialsymmetrische Gravitationsfelder* (Weyl'sche Klasse, Kerr-Lösung), für *ebene Wellen* und kosmologische Modelle, insbesondere für die *homogenen Kosmen,* bei denen alle Punkte in einem Zeitschnitt t = konst. gleichberechtigt sind. Bekannt sind außerdem zahlreiche Lösungen, deren physikalische Interpretation wenig befriedigt oder noch aussteht.

Verschiedenartige *Näherungsmethoden* wurden zur Untersuchung der Einsteinschen Feldgleichungen entwickelt (post-Newtonsche Korrekturen, linearisierte Feldgleichungen). Nach der *linearisierten Theorie,* die viele Gemeinsamkeiten mit der Maxwellschen Elektrodynamik hat und im Bereich schwacher Gravitationsfelder gilt, kann man die *Gravitationsstrahlung* berechnen, die von einer Quelle ausgesandt wird. Eine schlüssige Aussage über die Existenz von Gravitationsstrahlung in der nichtlinearen Theorie konnte bisher noch nicht getroffen werden.

Die Einsteinsche Theorie überwindet die Newtonsche Vorstellung von absolutem Raum und absoluter Zeit. Die Raum-Zeit ist nicht a priori gegeben, sondern wird gemäß $R_{ij} - (1/2)\, R g_{ij} = \varkappa T_{ij}$ durch die Materie bestimmt. Die Raum-Zeit-Geometrie stellt das Gravitationsfeld dar und ist somit selbst eine Form der Materie. Raum, Zeit und Materie bilden eine dialektische Einheit. Im Gegensatz zur Newtonschen Theorie ist die Einsteinsche Gravitationstheorie eine *Feldwirkungstheorie* (endliche Grenzgeschwindigkeit für die Ausbreitung von Wirkungen).

Die Einsteinsche Theorie erklärt zahlreiche Effekte mit hoher Genauigkeit und hat entscheidend zur Entwicklung der modernen Kosmologie (vgl. 3.5.) beigetragen.

Das unvermeidbare Auftreten von Singularitäten beim Sternkollaps und in kosmologischen Modellen deutet auf mögliche Grenzen der Theorie hin. Verschiedene nichteinsteinsche Gravitationstheorien wurden vorgeschlagen und teilweise bereits experimentell widerlegt. Eine künftige Theorie extrem starker Gravitationsfelder wird sicher auch Elemente der Quantentheorie beinhalten. Noch ist die Vereinigung von Gravitations- und Quantentheorie nicht gelungen.

3.4. Experimente zum Nachweis der Allgemeinen Relativitätstheorie (ART)

Sofort nach Aufstellung der ART begann eine intensive Suche nach meßbaren Abweichungen von der bisher üblichen klassischen Theorie. Trotz aller Bemühungen fand man aber nur die drei bekannten Effekte: Periheldrehung, Lichtablenkung am Sonnenrand sowie die Rotverschiebung im Schwerefeld von Sternen, die alle EINSTEIN bereits selbst angab. Doch waren auch hier die erzielten Meßergebnisse noch relativ ungenau und konnten Kritiker nicht immer überzeugen.
Genauso erwiesen sich die kosmologischen Folgerungen über die Flucht der Spiralnebel bei Annahme eines isotropen und homogenen Weltmodells und bei geeigneter Interpretation der Rotverschiebung in den Nebelspektren als qualitativ mit der Erfahrung in Übereinstimmung. Von Präzisionsmessungen konnte aber wegen der ungenauen Entfernungsbestimmung ferner Spiralnebel keine Rede sein.
Diese Situation blieb über ein halbes Jahrhundert so bestehen. Man konnte zwar erkennen, daß die Newtonsche Physik abgeändert werden mußte. Für die Einsteinsche Theorie sprachen aber weniger Meßwerte als Argumente der Geschlossenheit und mathematischen Schönheit und natürlich die vielen heuristischen Vorüberlegungen, die EINSTEIN zu seiner Theorie hinleiteten. Andererseits ist es auch verständlich, daß eine Vielzahl anderer nichteinsteinscher Gravitations- und Relativitätstheorien entstanden, z. B. die fünfdimensionalen Projektiven Relativitätstheorien, Theorien mit variabler Gravitationskonstante, Skalar-Tensor-Theorien, Bimetrische Theorien, Tetradentheorien, vierdimensionale Theorien mit Torsion (Einstein-Cartan) sowie speziell-relativistische und halbklassische Theorien.
Erst in den Jahren nach 1970 trat ein entscheidender Wandel ein. Durch Zunahme der Meßgenauigkeit speziell auch auf elektronischem Gebiet und durch den Einsatz der Mikroelektronik verbunden mit Fortschritten der Raketen- und Satellitentechnik gelang es, Effekte der ART EINSTEINS bis in den Prozent-, z. T. Promillebereich hinein nachzuweisen. Wesentlich war dabei, daß man sehr frequenzkonstante Sender bzw. entsprechend gute Reflektoren und programmgesteuerte Nachweisapparaturen an beliebige Stellen des näheren planetaren Raumes transportieren konnte.
Die erzielte Meßgenauigkeit reicht nicht nur aus, die Newtonsche Physik eindeutig zu widerlegen, sondern man kann auch zwischen den verschiedenen nichteinsteinschen Theorien unterscheiden. Viele von ihnen sind bereits widerlegt oder in ihrer Gültigkeit auf Parameterwerte in der Nähe der Einsteinschen Theorie eingeschränkt.
In allen diesen Theorien müßte man die gemessenen Effekte eigentlich gesondert rechnen. Da es sich aber um kleine Abweichungen von der Newtonschen Physik handelt, braucht man in der Reihenentwicklung nur um eine Ordnung über Newton hinauszugehen. Bis auf eine Ausnahme von Bedeutung (Einstein-Cartan-Theorie) kann man auf der Basis dieser linearen Näherung alle Theorien zu einem einheitlichen Formalismus vereinigen, in den zehn Parameter eingehen, deren Werte die verschiedenen (möglichen) Theorien unterscheiden. Das ist der sogenannte Parametrized Post Newtonian (PPN) Formalismus.
Die Experimente legen diese Parameter so fest, daß die Einsteinschen Werte überall innerhalb der Fehlergrenzen liegen. Letztere betragen vielfach nur wenige Prozent oder Promille. Man kann daher heute der Ansicht sein, daß die Einsteinschen ART in obigem Bereich genau richtig ist und erst der Einfluß der Quantisierung Änderungen erzwingt.

3.4.1. Übersicht über Experimente zur ART

Experimente oder Messungen zur Bestätigung der ART kann man danach einteilen, ob sie im Labor, im planetaren Raum, in unserer Galaxis oder im extragalaktischen Raum stattfinden.
Da die Gravitation die schwächste der vier Hauptwechselwirkungen ist, wird ihr Einfluß lokal außerordentlich klein, und es ist sehr schwierig, sie mit der für die ART nötigen Genauigkeit nachzuweisen. So ist bis 1981 keine signifikante Bestätigung der ART durch Laborexperimente gelungen. Das betrifft sowohl das Pound-Rebka Experiment (vgl. 3.4.7.) als auch die vergeblichen Versuche, Gravitationswellen nachzuweisen. Da man aber im Labor definierte Verhältnisse schaffen und alle Einzelheiten genau nachmessen kann, hofft man auf große Fortschritte auf diesem Gebiet.
Bereits im Planetensystem ist die Gravitation dominierend. Obwohl nach der ART bei großen Dimensionen (im Kosmos oder in Doppelsternsystemen größere Effekte auftreten, führt der mit der Entfernung zunehmende Mangel an Detailkenntnis und Meßgenauigkeit dazu, daß das Maximum der heute erzielbaren Genauigkeit im planetaren Raum liegt.

3.4.2. Beeinflussung elektromagnetischer Strahlung durch das Schwerefeld

Im Gravitationsfeld läuft ein Strahl einer elektromagnetischen Strahlungsquelle (Licht, Radar usw.) auf einer Nullgeodäten. Dadurch ergibt sich eine Richtungsablenkung und eine Laufzeitänderung bei Variation des Gravitationsfeldes. Auch die Frequenz wird beeinflußt. Damit die Effekte nachweisbar werden, müssen die Strahlen in der Nähe der Sonne verlaufen.
3.4.2.1. Ablenkung elektromagnetischer Strahlen im Schwerefeld der Sonne. Die Ablenkung eines Strahles

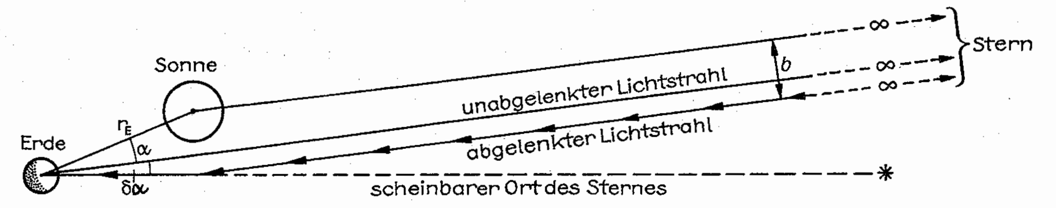


Abb. 3.4.2–1 Ablenkung elektromagnetischer Strahlung (Licht, Radar usw.) im Schwerefeld der Sonne

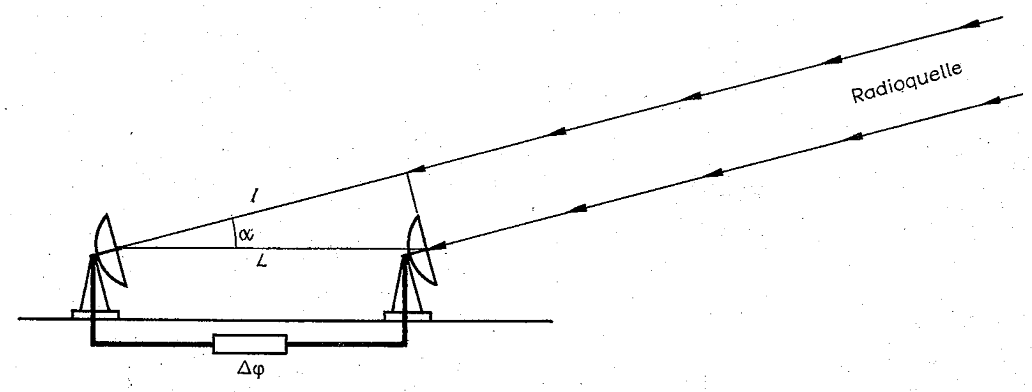


Abb. 3.4.2–2 Winkelmessung in der Radioastronomie durch Interferenz (Phasendifferenzmessung)

durch das Schwerefeld ergibt sich zu
$\delta\alpha = (1+\gamma)/2 \cdot 2\,GM_\odot/c^2 \cdot (1+\cos\alpha)/b$.
G ist die Newtonsche Gravitationskonstante, $M_\odot$ die Sonnenmasse, c die Lichtgeschwindigkeit, b der Abstand zwischen Strahl und Verbindungslinie Sonne – Stern in großer Entfernung, α ist der Winkel unter dem der Stern ohne Ablenkung erscheint, $\delta\alpha$ die relativistische Korrektur und γ der entsprechende PPN-Parameter. Tangiert der Strahl die Sonne, gilt $\delta\alpha = (1+\gamma)/2 \cdot 1''75$ (Abb. 3.4.2–1).
Die Werte von γ sind für die Theorien von Einstein: $\gamma = 1$, Jordan, Brans-Dicke $\gamma = (1+\omega)/(2+\omega)$.
ω ist ein freier Parameter in diesen Theorien, der durch die Messung von γ festgelegt wird. Für $\omega = \infty$ geht die Theorie in die Einsteinsche über. Diese Theorie kann also im PPN-Bereich nie widerlegt werden.
Die Experimente liefern aber sehr große ω-Werte und $\omega = \infty$ ist stets innerhalb der Fehlergrenzen.

Lichtablenkung am Sonnenrand. Das Licht von Sternen, die am Himmel dicht neben der Sonne stehen, wird in obiger Weise abgelenkt. Wegen der Überstrahlung kann man diesen Effekt nur bei totaler Sonnenfinsternis messen. Man fotografiert dieselbe Himmelsgegend mit demselben Instrument einmal während die Sonne weit entfernt ist und dann während der Finsternis und vergleicht beide Platten. Schwierigkeiten kommen von der Sonnenkorona, die die Strahlen passieren müssen und deren Brechzahl man abschätzen muß. Durch die Abnahme der Sonneneinstrahlung während der Finsternis wird die gesamte Atmosphäre instabil. Die auftretenden Turbulenzen begrenzen die Genauigkeit der Abbildung. Die Instrumente und Linsen beginnen sich durch Abkühlung zusammenzuziehen, was man ausgleichen muß. Trotz aller Korrekturen bleiben bei den besten Messungen die Fehler über 10 %.

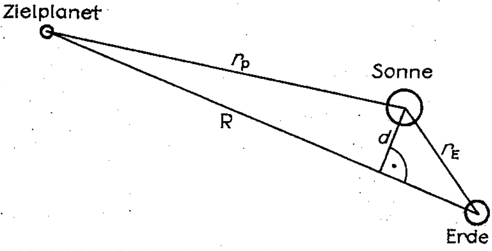


Abb. 3.4.2–3 Radarechomethode

CROMMELIN, DAVIDSON (Sobral)	1919	1″98 ± 0″18
EDDINGTON, COTTINGHAM (Principe)	1919	1″61 ± 0″45
CAMPBELL, TRÜMPLER (Wallal)	1922	1″72 ± 0″15
DODWELL, DAVIDSON (Cordillo Downs)	1922	1″77
FREUNDLICH, KLÜBER, BRUNN (Takuya)	1929	2″24
MICHAILOW (Kuybyshevka)	1936	2″73 ± 0″31
VAN BIESBROECK (Bocayura)	1947	2″01 ± 0″27
VAN BIESBROECK (Khartum)	1952	1″70 ± 0″10

Ablenkung der Strahlung von Radioquellen. Die Radioquelle 3C279 liegt so nahe der Ekliptik, daß die Strahlablenkung bei Vorbeigang der Sonne gut meßbar wird. Da die Radioparabolantennen eine schlechte Winkelauflösung haben, werden Interferenzschaltungen mehrerer solcher Antennen verwendet mit Basislängen von wenigen bis einigen tausend Kilometern. Gemessen wird über viele Tage die zeitliche Änderung des Winkelabstandes zu anderen Radioquellen.
Aus der Wegdifferenz l (Abb. 3.4.2–2) ergibt sich eine Phasendifferenz $\varphi = 2\pi l/\lambda = \omega l/c$, die kontinuierlich gemessen wird. Daraus berechnet sich der Winkel zu $\cos\alpha = c\varphi/\omega L$. Eine Sonnenfinsternis ist nicht nötig. Durch Messung bei verschiedenen Abständen und auf verschiedenen Wellenlängen kann man den Einfluß der Korona abschätzen. Schwierigkeiten entstehen auch aufgrund der eventuell geometrisch weit getrennten Wege durch die Erdatmosphäre mit verschiedenen optischen Weglängen, zu deren Korrektur (wegen Luftdruck und -feuchtigkeit) zusätzliche Messungen nötig sind. Bei sehr entfernten Antennen werden die Signale auf Bänder aufgenommen, auf die durch sehr frequenzkonstante Wasserstoffmaser Zeitmarken aufgetragen werden.
Die ersten Messungen stammen von 1970. Die Genauigkeit beträgt zur Zeit ≈1 % des Einsteinschen Wertes.
1974 COUNSELMAN, SHAPIRO et al. ($L = 845$ km)
$(1+\gamma)/2 = 0{,}99 \pm 0{,}03$
1976 FOMALANT, SRAMEK ($L = 35$ km)
$(1+\gamma)/2 = 1{,}015 \pm 0{,}011$

3.4.2.2. Laufzeitänderung von Radarechos im Schwerefeld der Sonne. Radarsignale kann man zu den benachbarten Planeten (Merkur, Venus, Mars) oder Satelliten senden und die Echos auffangen. Die Laufzeit hängt nicht nur von der Entfernung ab, sondern auch vom durchquerten Gravitationsfeld. Durch die Sonne ergibt sich folgende Laufzeitänderung (Abb. 3.4.2–3):

$$\Delta\tau = \frac{1+\gamma}{2}\cdot\frac{2}{c}\cdot\frac{2GM_\odot}{c^2}\ln\frac{r_E + r_P + R}{r_E + r_P - R}.$$

Wegen der Kugelgestalt und der Gebirge ($h \approx 10$ km) sind die Echos von Planeten nicht ideal, auch wenn man die Oberflächenstruktur im Moment der Reflexion kennt. Besser sind die aktiven Targets künstlicher Satelliten, die im Moment der Signalankunft ein Signal zurücksenden, obwohl ihre Bahn nicht so stabil und auch nicht so genau bestimmbar ist. Die größte Genauigkeit (1 ‰) erzielt man durch Satelliten, die auf Planeten gelandet sind (Viking-Lander auf Mars, wobei allerdings auch die Orbiter zur Signalrücksendung benutzt werden). Als Korrektur muß noch die Laufzeitänderung in der Korona ermittelt werden. Dazu werden mehrere Signale verschiedener Frequenz gleichzeitig durch die Korona gesendet. Während $\Delta\tau$ frequenzunabhängig ist, gilt $\tau_{\text{corona}} = \frac{(300)}{f^2 \cdot d}\,\mu\text{sec}$. f ist die Frequenz, d der Abstand aus Abb. 3.4.2–3. Der Faktor (300) ist nur ein Mittelwert. Er kann sich kurzzeitig stark ändern und muß daher im Moment der Messung bestimmt werden.
Die Resultate sind:

			$(1+\gamma)/2 =$
1968/72	Venus, Merkur	SHAPIRO	$1{,}02 \pm 0{,}04$
1971/76	Mariner 6,7	ANDERSON	$1{,}00 \pm 0{,}03$
ab 1971	Mariner 9	SHAPIRO et al.	$1{,}00 \pm 0{,}02$
1976	Viking auf Mars	SHAPIRO, REASENBERG	$1{,}000 \pm 0{,}001$

3.4.3. Periheldrehung

Zur Erklärung der Differenz zwischen gemessener Periheldrehung und dem aus der üblichen nichtrelativistischen Störungstheorie folgenden Wert (bei Merkur 43,″11/Jahrhundert) muß man die ART und das Quadrupolmoment der Sonne verantwortlich machen. Beide Effekte zusammen ergeben eine Periheldrehung von

$$\delta\varphi = \frac{6\pi(GM_\odot/c^2)}{a\cdot(1-\varepsilon^2)} \cdot \left\{\frac{2-\beta+2\gamma}{3} + J_2\frac{R_\odot^2/(GM_\odot/c^2)}{2a\cdot(1-\varepsilon^2)}\right\}.$$

Dabei ist a die große Halbachse und ε die numerische Exzentrizität, $M_\odot$ bzw. $R_\odot$ sind Masse und Radius der Sonne. J_2 ist ihr (unbekanntes) Quadrupolmoment und β PPN-Parameter ($\beta_{\text{Einstein}} = 1$). Durch optische und Radarmessungen ergeben sich für die Klammern (SHAPIRO et al. 1976):

$\{\ \}_{\text{Merkur}} = 1{,}003 \pm 0{,}005$, $\{\ \}_{\text{Mars}} = 1{,}07 \pm 0{,}10$.

Durch Ausmessung mehrerer Bahnen (z. B. Merkur und Mars) erhält man mehrere Gleichungen für die zwei Unbekannten $(2-\beta+2\gamma)/3$ und J_2. Jedoch sind alle Meßwerte außer für Merkur relativ ungenau. Das überträgt sich auf J_2 und β (wobei aber β_{Einstein} innerhalb der Meßgrenzen liegt):

$$J_2 \approx -3\cdot 10^{-5} \pm 5\cdot 10^{-5}, \quad \beta = 1{,}0 + \begin{cases} +0{,}2 \\ -0{,}7 \end{cases}.$$

Geplante Messungen an geeigneten Satelliten in stark elliptischen Bahnen sollen genauere Werte liefern. Die Absolutwerte der Periheldrehung (ohne Quadrupolmoment berechnet) sind die folgenden:

Periheldrehung/Jahrhundert	ART	gemessen
Merkur	43,″03	43,″11
Venus	8,″6	8,″4
Erde	3,″8	5,″0

3.4.4. Lunar-Laser-Ranging

Durch amerikanische und sowjetische Raketen sind im ganzen fünf Scheiben mit Prismenreflektoren auf den Mond postiert worden. Diese reflektieren auftreffendes Licht unabhängig vom Einfallswinkel genau in die Herkunftsrichtung zurück (Abb. 3.4.4–1). Ein normaler Spiegel würde das höchstens für einen Moment leisten und wäre dann durch Rotation, Libration, Erschütterungen usw. dejustiert. Durch Messung der Laufzeit von Laserimpulsen kann man die Entfernung zwischen Erde und Mond bis auf Dezimeter genau bestimmen. Durch Benutzung mehrerer Spiegel erhält man die Lage des Mondmittelpunktes mit solcher Genauigkeit, daß auch die relativistischen Bahnkorrekturen nachgemessen werden können (Nordtvedt-Effekt).
Die Bewegungsgleichung des Mondes lautet in diesem Fall

$$\underbrace{\ddot{\boldsymbol{r}} + \frac{G(M_♁ + M_☾)}{r^3}\cdot \boldsymbol{r}}_{\text{Zweikörperproblem Erde–Mond}}$$

$$= \underbrace{-GM_\odot\left\{\frac{\boldsymbol{r}_☾}{r_☾^3} - \frac{\boldsymbol{r}_♁}{r_♁^3}\right\}}_{\text{klassisches Störglied}} + \underbrace{\eta\, GM_\odot(\Delta_♁ - \Delta_☾)\frac{\boldsymbol{r}_♁}{r_♁^3}}_{\text{relativistische Korrektur}}.$$

Dabei ist $\Delta = M_g/M$ = gravitative Selbstenergie/Ruhenergie. Für Körper von 1 m Abmessungen ist $\Delta \approx 10^{-23}$, sonst gilt $\Delta \sim \varrho R^2$. η ist PPN-Parameter, $\eta_{\text{Einstein}} = 0$. Seine Bedeutung hängt mit der Gleichheit von schwerer und träger Masse zusammen $M_{\text{schwer}}/M_{\text{träge}} = 1 + \eta\cdot\Delta$.
Durch Lösung obiger Differentialgleichung mittels Computer erhält man $\eta = 0 \pm 0{,}015$. Daraus folgt für den Mond $M_s/M_{\text{träge}} = 1 \pm 10^{-13}$.

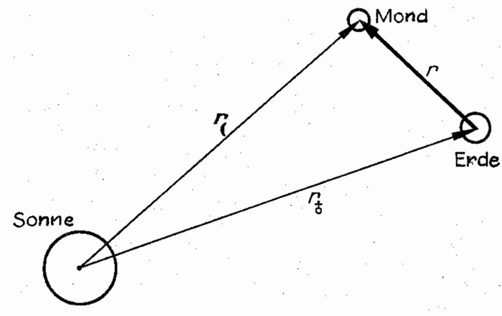


Abb. 3.4.4–1 Lunar-Laser-Ranging

3.4.5. Äquivalenzprinzip

Die Ergebnisse des Nordtvedt-Effektes beweisen das Äquivalenzprinzip, das auf der Gleichheit von schwerer und träger Masse beruht. Es war bisher bereits mit großer Genauigkeit durch Drehschwingungen getestet.

	$\varepsilon = M_s/M_{\text{träge}} - 1$
EÖTVÖS	$\lvert\varepsilon\rvert < 5\cdot 10^{-9}$
ROLL-KROTKOV	$\lvert\varepsilon\rvert < 3\cdot 10^{-11}$
BRAGINSKY	$\lvert\varepsilon\rvert < 3\cdot 10^{-12}$
Nordtvedt-Effekt (1 kg)	$\lvert\varepsilon\rvert < 10^{-26}$ (Umrechnung auf 1 kg Masse).

3.4.6. Messung der Änderung der Gravitationskonstanten

Eine Änderung der Gravitationskonstante würde zu einer Änderung der großen Halbachse a und der Umlaufszeit P führen. $\dot{G}/G = -\dot{a}/a = -\dot{P}/2\,P$. Durch Vergleich alter Messungen zeigte VAN FLANDERN $\dot{G}/G \leq 10^{-11}$/Jahr, während REASENBERG und SHAPIRO durch Radarmessungen den Wert $|\dot{G}/G| \leqq 1{,}5 \cdot 10^{-10}$/Jahr fanden.

3.4.7. Gravitationsrotverschiebung

Läuft eine Lichtwelle in einem zeitunabhängigen Gravitationsfeld von einem Punkt P_0 nach P, so verändert sich die Frequenz von ν_0 auf ν, $\nu\sqrt{-g_{44}(P)} = \nu_0\sqrt{-g_{44}(P_0)}$.
Im kugelsymmetrischen Feld einer Masse M gilt

$$-g_{44} = 1 - 2\frac{GM/c^2}{r} + 2\beta\left(\frac{GM/c^2}{r}\right)^2.$$

Auf der Sonne ist der Effekt vorhanden, aber durch Strömungs-Doppler-Verschiebung überlagert. Für den Sirius-Begleiter (weißer Zwerg) kann man den theoretischen Wert nachmessen (Fehler 15 %). Genauer sind die Messungen von POUND und REBKA mittels Mößbauer-Effekt, wo man die Frequenzverschiebung im Schwerefeld der Erde auf einer Höhe von nur 20 m mit ≈1 % Fehler nachweisen kann. Analoge Versuche wurden mit sehr frequenzkonstanten Uhren in Flugzeugen (10 km Höhe) und Raketen (10 000 km) wiederholt. Das beste Resultat lieferte ein Raketenaufstieg auf 10 000 km Höhe mit einer Wasserstoffmaseruhr an Bord, wobei der Fehler nur 0,014 % des theoretischen Wertes betrug. Allerdings reicht auch hier die Genauigkeit noch nicht aus, das Glied mit β nachzuweisen, was erst Aufgabe eines geplanten Sonnensatelliten sein wird (Abb. 3.4.7-1).

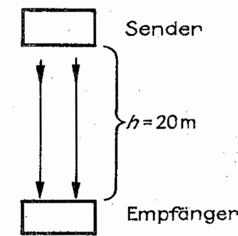


Abb. 3.4.7-1
Pound-Rebka-Experiment

3.4.8. Nachweis von Gravitationswellen

In linearer Näherung gleicht die Einsteinsche Theorie in gewisser Hinsicht der Elektrodynamik. Es ist hier sehr einfach einzusehen, daß es Wellen gibt (Quadrupolstrahlung), wie sie sich ausbreiten und welche Ausstrahlungsbedingungen (retardierte Potentiale) man fordern muß. In der allgemeinen Theorie sind diese Fragen aber wegen zu großer Schwierigkeiten immer noch unbeantwortet. Auch die Konvergenz entsprechender Reihen ist nicht bewiesen, obwohl man auf diesem Gebiet schon Fortschritte erzielt hat. So ist es nicht abwegig, daß gelegentlich Meinungen geäußert wurden, Gravitationswellen im strengen Sinne würden überhaupt nicht existieren.
Insofern wäre ein experimenteller Nachweis von großer Wichtigkeit. Rechnungen möglicher Gravitationsstrahlungsquellen im Weltall (Supernovaexplosionen, Kollaps, Durchdringung zweier Sternhaufen, Doppelsterne) zeigen, daß maximal Amplituden von $h \lesssim 10^{-19}$ zu erwarten sind.

3.4.8.1. Messung von Gravitationswellen im Labor. Trotz vieler Bemühungen sind Gravitationswellen noch nicht nachgewiesen worden. Die am weitesten ausgearbeitete Idee stammt von J. WEBER. Die Gravitationswellen fallen auf einen Festkörper mit möglichst gleicher Eigenfrequenz. Dessen Schwingungen werden kapazitiv ausgekoppelt und elektronisch verstärkt. Anfangs verwendete man Aluminiumzylinder, später zur Erhöhung des Gütefaktors Saphireinkristalle, die speziell aus der Schmelze gezogen werden. Die gesamte Anlage wird in flüssigem Helium gekühlt, um das Rauschen zu verkleinern. BRAGINSKY hat durch besonderes Schleifen des Saphireinkristalls noch einen Hohlraumresonator mit Wänden aus supraleitendem Niobium integriert, über den die elektronische Auskopplung des Signals erfolgt. Erreichen kann man heute $h = \Delta\, l/l = 3 \cdot 10^{-17}$ (WEBER, BRAGINSKY, DOUGLASS, TYSON). Bei Steigerung der Empfindlichkeit kommt man in den Bereich der Quantentheorie und muß deren Besonderheit berücksichtigen, um nicht schnell auf Nachweisschranken zu stoßen. Wesentlich scheint hierbei die Methode der unzerstörten Quantenmessung (quantum nondemolition measurement) zu sein.

3.4.8.2. Messung an Doppelsternsystemen. Bessere Resultate schien die Untersuchung des spektroskopischen Doppelsternsystems 1912 + 16 zu bieten, dessen eine Komponente ein Pulsar ist. Im Rahmen des üblichen Pulsarmodells kann man aus Pulsdauer und deren Modulation, aus Frequenz und Frequenzverschiebung sowie aus Amplituden- und Polarisationsmessung die Bahnelemente des Doppelsterns berechnen, wobei die allgemein-relativistische Mechanik mit einer entsprechend großen Periastrondrehung zugrunde gelegt wird. Aus den berechneten Bahnelementen folgt auch das Quadrupolmoment und die Gravitationswellenausstrahlung je Zeiteinheit sowie die damit verbundene Abnahme der großen Halbachse. Die Genauigkeit der Messungen ist so groß, daß nach mehrjährigen Beobachtungen festgestellt werden kann, ob die große Halbachse entsprechend den linearen Ausstrahlungsformeln abnimmt oder nicht. Allerdings ist Voraussetzung, daß der zweite Stern nahezu punktförmig ist und nicht so groß, daß die Bahn des Pulsars durch Reibung der äußersten Hüllen beeinflußt wird, oder daß Überströmung von Materie stattfindet, was andere Mechanismen zur Erklärung der experimentellen Daten fordern würde. Neuere Analysen des Spektrums sprechen für einen im Vergleich zur Pulsarbahn großen Stern (Heliumstern), so daß man zur Zeit leider keine klare Aussage über Ausstrahlung machen kann.

3.4.9. Kosmologische Tests der ART

Wie in der Einleitung bereits erwähnt, besteht sichtbar Übereinstimmung zwischen einem nach der ART berechneten Friedman-Kosmos und der gemessenen Entfernungsabhängigkeit der Rotverschiebung von Nebellinien. Von Präzisionsmessung kann aber keine Rede sein. Schon der „Beschleunigungsparameter“ (der Expansion) hat Fehlergrenzen von ≈100 %. Aussagekräftiger ist dagegen die gemessene Isotropie und Intensität der elektromagnetischen Hintergrundstrahlung (vgl. 3.5.4.2., auch über Grenzen). Durch Rechnungen über den Frühzustand des Kosmos kann man zeigen, daß eine solche im wesentlichen isotrope Strahlung auftreten muß und wie heute ihre Intensitätsverteilung ist. Ein weiteres Resultat dieser Rechnung, die heu-

tige Häufigkeit der Elemente (vor allem Wasserstoff und Helium) läßt sich ebenfalls experimentell verifizieren.

3.5. Allgemeinrelativistische Kosmologie

3.5.1. Gravitationstheorie – Grundlage der Kosmologie

Die Struktur und Entwicklung von physikalischen Systemen, damit auch des Weltalls (griechisch: kosmos = Welt), wird durch die Wechselwirkung (Ww) verschiedener Teile des Systems untereinander sowie zwischen System und Umgebung bestimmt. Abgesehen von extremen Verhältnissen (vgl. 3.6.), kann man die bisher bekannten Ww in vier Gruppen einteilen: die starke Ww (hierzu gehören die *Kernkräfte*), die elektromagnetische (z. B. Coulomb-Kraft), die *schwache* (Ww unter Beteiligung von *Neutrinos*, z. B. *β-Zerfall*) und die gravitative (Massenanziehung). In kosmischen Dimensionen spielt die gravitative Ww die dominierende Rolle. Sie hat zwar die geringste Kraftkonstante, aber die größte effektive Reichweite. Die Quellen der elektromagnetischen Kräfte, die Ladungen, können sich gegenseitig kompensieren (positive – negative Ladungen). Diese Kompensation gibt es für die Gravitation wegen des Fehlens negativer Massen nicht. Die Grundannahme jeder Kosmologie ist daher die folgende: *Die Bewegung der Materie wird im wesentlichen durch das Gravitationsfeld bestimmt.* Man erwartet also zu jeder Gravitationstheorie eine entsprechende Kosmologie. Die auf der *Newtonschen Gravitationstheorie* basierende Kosmologie enthält ernste Widersprüche. Ein wirklicher Fortschritt beim Aufbau einer wissenschaftlichen Kosmologie gelang erst nach Schaffung der allgemeinrelativistischen Gravitationstheorie.

3.5.2. Kosmologisches Prinzip

Zur Bestimmung der Dynamik des Weltalls sind die Einsteinschen Feldgleichungen der Gravitation zu lösen. Dies geschieht bei allen Weltmodellen unter speziellen Zusatzannahmen über die Beschaffenheit und Verteilung der Materie im Kosmos. Dazu geht man vom heutigen Zustand aus: Die kosmische Materie wird durch ein Gas ersetzt, in dem die Galaxien die Rolle der Atome übernehmen. Über Regionen gemittelt, die groß gegen den Galaxienabstand sind, ist die Dichte des Gases im der Beobachtung zugänglichen Raum näherungsweise konstant (Homogenität) und hängt nicht von der Beobachtungsrichtung ab (Isotropie). Durch Verallgemeinerung der an der Dichte festgestellten Eigenschaften ergibt sich das *kosmologische Prinzip* (K. P.): Das *gesamte* Weltall ist bezüglich *aller* Größen, die seinen Zustand charakterisieren, räumlich homogen und isotrop.

3.5.3. Friedman-Modelle

Bei Gültigkeit des K. P. kann die vierdimensionale Riemannsche Raumzeit in einen dreidimensionalen Ortsraum konstanter Krümmung und eine universelle Zeit aufgespalten werden. Das Linienelement hat die Form *(Robertson-Walker-Metrik)*

$$ds^2 = K^2(t)\,[dr^2/(1 - kr^2) + r^2(d\vartheta^2 + \sin^2\vartheta\, d\varphi^2)] - c^2 dt^2$$
$$\equiv K^2(t)\, ds^{(3)2} - c^2 dt^2.$$

Der Parameter k kann drei Werte annehmen:

$$k = \begin{cases} -1 & \text{offener (negativ) gekrümmter Ortsraum (hyperbolische Geometrie)} \\ 0 & \text{offener flacher Ortsraum (parabolische Geometrie),} \\ 1 & \text{geschlossener positiv gekrümmter Ortsraum (sphärische Geometrie)} \end{cases}$$

Offen heißt ein Raum mit unendlich großem Volumen. $K^2 ds^{(3)}$ ist das Abstandsquadrat zweier infinitesimal benachbarter Punkte im Ortsraum. Der Faktor $K(t)$ zeigt an, daß der Abstand zeitabhängig ist. Gilt $K(t_2) > K(t_1)$ im Zeitintervall $[t_1, t_2]$, so wächst der Abstand der Punkte. Man nennt diesen Vorgang *Expansion* des Ortsraumes (*Kontraktion* für $K(t_2) < K(t_1)$). Die Expansion bzw. Kontraktion eines positiv gekrümmten Raumes kann man sich leicht anhand eines gleichmäßig mit punktförmigen Farbtupfen versehenen ideal runden Luftballons veranschaulichen. Die Oberfläche des Luftballons repräsentiert dabei das endliche Volumen dieses zweidimensionalen Analogons des dreidimensionalen sphärischen Ortsraumes. Beim Aufblasen des Ballons entfernen sich sämtliche Tupfen voneinander (Expansion). Entscheidend für das zeitliche Verhalten des Abstandes ist die Funktion $K(t)$. Im zweidimensionalen Fall ist $K(t)$ identisch mit dem Radius der Kugeloberfläche. Für den dreidimensionalen Raum konstanter Krümmung wird $K(t)$ Weltradius des kosmologischen Modells genannt. Das zeitliche Verhalten des Weltradius bestimmen die Einstein-Gleichungen. Sie verknüpfen bei Annahme des kosmologischen Prinzipes in recht einfacher Form $K(t)$ mit dem Druck $p(t)$ und der Energiedichte $\varepsilon(t)$ der kosmischen Materie ($\dot{K} = dK/dt$):

$$\dot{K}^2/c^2 + k = \varkappa\varepsilon K^2/3, \quad (K^3\varepsilon)^{\cdot} + p(K^3)^{\cdot} = 0.$$

Die durch diese Differentialgleichungen (Friedman-Gleichungen) beschriebenen Weltmodelle heißen Friedman-Modelle (1922). Zur Ermittlung der Lösungen $K(t)$, $p(t)$, $\varepsilon(t)$ muß der Zustand der kosmischen Materie in Form einer Zustandsgleichung $p = p(\varepsilon)$ näher charakterisiert werden. Unter recht allgemeinen Annahmen über den Zustand der kosmischen Materie ergibt sich qualitativ die Zeitabhängigkeit des Weltradius $K(t)$ (Abb. 3.5.3-1). Die wesentliche Eigenschaft der Friedman-Weltmodelle liegt darin, daß

Abb. 3.5.3-1 Zeitabhängigkeit des Weltradius für die möglichen Friedman-Modelle; t_0 = Zeitpunkt des gegenwärtigen Zustandes

sie nicht statisch sind, entweder sie expandieren oder kontrahieren. Wird die Expansion vom heutigen Zeitpunkt t_0 zurück zu früheren Zeiten verfolgt, so kommt man zu einem Zustand mit verschwindendem Weltradius ($K = 0$), von dem die Expansion ausging. Diesen Punkt nannte G. GAMOW „big bang" (engl. „Urknall"). Die Zeitspanne vom Urknall bis zum heutigen Zeitpunkt heißt Weltalter t_0. Man sagt auch, daß zu Zeitpunkten, in denen alle Abstände auf Null zusammengeschrumpft sind ($K = 0 \rightarrow ds^2 = 0$), eine Singularität vorliegt. Ob solche Singularitäten überhaupt vermeidbar sind, ist noch nicht völlig geklärt. Der Zustand der Materie nahe einer Singularität dürfte jedoch mit den bisher bekannten Naturgesetzen kaum beschreibbar sein. Über die Brauchbarkeit des Friedman-Modells als theoretische Grundlage der realen Welt sollen neben der Verteilung der Galaxien weitere Beobachtungsergebnisse entscheiden.

3.5.4. Beobachtungsergebnisse

3.5.4.1. Hubble-Expansion. E. HUBBLE entdeckte 1929 an entfernten Galaxien eine isotrope Rotverschiebung $z = \Delta\lambda/\lambda$ der Spektrallinien. Diese Rotverschiebung kann durch ein zeitlich expandierendes Universum widerspruchsfrei erklärt werden. Die strenge Rechnung liefert folgende Entfernungs-Rotverschiebungsrelation: $d_L = [z + 1/2\,(1 - q)\,z^2 + \ldots]\,c/H$. Für die Hubble-„Konstante" ergibt sich $H(t) = \dot{K}(t)/K(t) > 0$ (bei Expansion). $q(t) = -\ddot{K}K/\dot{K}^2$ ist der sogenannte Beschleunigungsparameter ($q \geq 0$ generell). Nimmt man an, daß das Universum unbeschleunigt expandiert, folgt $t_0 = 1/H_0$ als Weltalter. Diese Abschätzung von t_0 ist recht gut, da das wirkliche Weltalter nur schwach von q abhängt.

3.5.4.2. Beobachtungen zur Geometrie des Modells (Entscheidung über k). Der Zustand der kosmischen Materie zum heutigen Zeitpunkt wird durch $p \approx 0$ (d. h. $\varepsilon_0 \approx \varrho_0 c^2$) gut beschrieben. Nach Einführung von H, q liefern die Friedman-Gleichungen für

$$t = t_0\colon\ \varkappa\varrho_0 c^2 = 6\,H_0^2 q_0/c^2,\ k = H_0^2 K_0^2 (2\,q_0 - 1)/c^2.$$

Es hängt von der Ruhmassendichte ϱ_0 ab, ob der Ortsraum offen oder geschlossen ist. Die Krümmung des Ortsraumes vergrößert sich mit steigender Massendichte. Den Grenzfall zwischen offenem und geschlossenem Modell bildet $k = 0$, d. h. $q_0 = 1/2$ (parabolische Geometrie), die entsprechende Ruhmassendichte heißt die *kritische Massendichte* $\varrho_{krit} = 3\,H_0^2/(\varkappa c^4)$. Die Messung von H_0 ist noch recht ungenau. Wahrscheinlich erscheint der Wert $H_0 \approx 55$ km/s · Mpc ($H_0^{-1} \approx 1{,}8 \cdot 10^{10}$ Jahre = Weltalter). Damit ergibt sich $\varrho_{krit} \approx 6 \cdot 10^{-30}\,\mathrm{g\,cm^{-3}}$. Rotverschiebungsmessungen liefern eine derzeitige Dichte $\varrho_0 \approx 1{,}2 \cdot 10^{-29}\,\mathrm{g\,cm^{-3}}$ ($q_0 \approx 1$), Galaxienzählungen dagegen $\varrho_0 \approx 1{,}2 \cdot 10^{-31}\,\mathrm{g\,cm^{-3}}$. Der Nachweis bisher unentdeckter Massen würde den Widerspruch zwischen beiden Meßmethoden mindern. Die Mehrheit der Astrophysiker geht davon aus, daß beide Messungen fehlerhaft sind und rechnet mit $\varrho_0 \approx 10^{-30}\,\mathrm{g\,cm^{-3}}$, danach wäre das Universum offen. Aus der Existenz von Neutrinos mit Ruhmasse würde eindeutig ein geschlossenes Universum folgen.

3.5.4.3. Kosmologische Hintergrundstrahlung. War bereits die Rotverschiebung eine wesentliche Stütze für die Anwendbarkeit des Friedman-Modells der Einstein-Gleichungen, die Entdeckung der kosmologischen Hintergrundstrahlung (1965) hat dieses Modell geradezu favorisiert. PENZIAS und WILSON fanden bei Untersuchungen eines Radioteleskopes bei einer Wellenlänge von $\approx 7{,}35$ cm einen nicht zu eliminierenden Störpegel. Es zeigte sich, daß die Radioantenne eine aus dem Kosmos isotrop einfallende elektromagnetische Strahlung registrierte, die die Energieverteilung einer Planckschen Hohlraumstrahlung mit einer Temperatur von 2,7 K (3-K-Hintergrundstrahlung, Mikrowellen-Hintergrundstrahlung) aufwies. Bis etwa Anfang der siebziger Jahre hatte man die Strahlung im Wellenlängenbereich von 0,1 cm bis 74 cm zum Teil durch Infrarotmessungen mit Hilfe von in Ballons bzw. in U2-Flugzeugen (Flughöhe $\approx$ 20 km) installierten Meßanlagen, zum Teil durch Untersuchung der Rotationsbanden der CN-Moleküle von interstellaren Gaswolken (z. B. im Sternbild Schlangenträger) eingehend erforscht und mit einem Fehler von weniger als 1 % keinerlei Richtungsabhängigkeit (Anisotropie) und Polarisation nachgewiesen. Die Strahlungsmeßdaten beziehen sich jedoch auf das irdische Bezugssystem, das sich bekanntlich gegen die kosmische Hauptmasse bewegt ($v_{Erde} \approx 300$ km/s). Diese Bewegung müßte sich in einer Anisotropie unpolarisierter Strahlung von $\approx 0{,}1$ % äußern. Neuere Messungen (1976 bis 1979) haben nicht nur die Anisotropie aufgrund der Bewegung der Erde durch das Strahlungsfeld (Dipolanisotropie) bestätigt, sondern liefern auch eine Quadrupolanisotropie (Größenordnung $\approx$ 1/3 Dipolanisotropie). Die Deutung der Quadrupolanisotropie ist noch umstritten. Nach Modellrechnungen sind gewisse homogene anisotrope Störungen der Friedman-Metrik verantwortlich, die zeitlich anwachsen aber erst in 10^{13} bis 10^{18} Jahren (je nach Massendichte des Universums) die großräumige Struktur so verändert haben, daß das derzeit dem Friedman-Modell sehr ähnliche Universum dann merklich von der Friedman-Geometrie abweicht. Diese Deutung stünde mit den Ergebnissen von COLLINS/HAWKING, nach denen fast alle Friedman-Modelle (bis auf das Einstein-de-Sitter (EdS)-Modell $\mathrel{\hat{=}} k = 0$) instabil sind, im Einklang. Die 3-K-Strahlung liefert weitere Informationen über die Entwicklung des Universums.

3.5.5. Thermodynamische Betrachtungen

Durch Einführung eines speziellen Volumens V (Volumen eines beliebigen Teiles des gesamten Kosmos) liefert die Gleichung $(K^3\varepsilon)^{\boldsymbol{\cdot}} + p(K^3)^{\boldsymbol{\cdot}} = 0$ eine wichtige thermodynamische Aussage über das Gesamtsystem Kosmos: $\dot{E} + p\dot{V} = 0$ ($E = \varepsilon V$, $V = K^3$). Diese Gleichung ist der sogenannte 1. Hauptsatz eines thermodynamischen Systems mit zwei Freiheitsgraden (z. B. eines Gases) für einen adiabatischen Prozeß (kein Wärmeaustausch zwischen System und Umgebung). Man nimmt an, daß im Kosmos verschiedene Materieformen (Stoff, bestehend aus einzelnen chemischen Komponenten; freie Elementarteilchen bzw. Atomkerne; Strahlung) existieren, die unter bestimmten Bedingungen auf unterschiedlichste Weise wechselwirken. Entscheidende Information gibt der heutige Zustand. Zunächst existiert ein Gas (Galaxiengas) mit vernachlässigbarem Druck, außerdem die 3-K-Hintergrundstrahlung ($T_\gamma \approx 3$ K). Diese beiden Komponenten wechselwirken derzeit nicht, verhalten sich also wie separate Systeme. In einer früheren Entwicklungsphase ist auch die thermische Energie des Gases zu berücksichtigen, die in E_R enthalten sei. Es gilt: $E = E_G + E_\gamma + E_R$, $p = p_{Gas}(\approx 0) + p_\gamma + p_R$ (E_G = Ruhenergie des Gases, E_γ = Energie der elektromagnetischen Hintergrundstrahlung, E_R = hier nicht näher charakterisierter

Tab. 3.5.6–1 Temperaturgeschichte des Universums

T in Kelvin	t (nach dem Urknall)	Komponenten	Bemerkungen
$> 10^{12}$	$<0{,}01$ s	γ, $\mu^{\pm}$, $e^{\pm}$, ν, $\bar{\nu}$ (geringe Mengen p, n)	Komponenten sind alle im thermischen Gleichgewicht; ν, $\bar{\nu}$ – Neutrinos, Antineutrinos; Zahl der Protonen N_p = Zahl der Neutronen N_n
10^{11}	0,01 s	γ, $e^{\pm}$, ν, $\bar{\nu}$; $\mu^{\pm}$ durch Annihilation fast vernichtet	$N_p > N_n$, wegen größerer Neutronenmasse
10^{10}	1 s	γ, p, n, $e^{\pm}$; ν, $\bar{\nu}$	im thermischen Gleichgewicht; aus Gleichgewicht ausgekoppelt; Kosmos wird für Neutrinos durchsichtig, aber $T_\gamma = T = T_{\bar{\nu}}$; $N_p : N_n = 5:1$
$3 \cdot 10^9$	14 s	γ, p, n, e^-, $(e^{\pm})$; ν, $\bar{\nu}$	$N_{e^-} = N_p$, da Gesamtladung Null; durch Zerstrahlung von $e^{\pm} \rightarrow T_\gamma = T > T\nu$
$1{,}2 \cdot 10^9$	3 min	γ, p, n, e^-; ν, $\bar{\nu}$; $e^{\pm}$ fast völlig annihiliert	durch β^--Zerfall des Neutrons wird $N_p : N_n > 5:1$
10^9	3 min 45 s	γ, p, H_1^2, He^3, He^4, e^-; ν, $\bar{\nu}$	über Kernfusion durch Zweierstöße entstehen ≈ 26 % He, den Hauptanteil der kosmischen Ruhmasse stellen die Protonen
$4 \cdot 10^3$	$3 \cdot 10^5$ a	γ, H, He; ν, $\bar{\nu}$	Kombination von e^- und Atomkernen zu neutralen Atomen; Kosmos wird für elektromagnetische Strahlung durchsichtig (Entkopplung), aus dieser Zeit Hintergrundstrahlung mit $T_\gamma(t)$
Mußte man ab $t = 14$ s getrennte Temperaturen für Neutrinos und Rest angeben, so sind nun drei Temperaturen $(T_\nu, T_\gamma, T_{Rest})$ zu kennzeichnen, darauf wird hier verzichtet.			
	10^8 a	Entstehung von Protogalaxien, in Galaxien können sich schwere Elemente bilden; γ; ν, $\bar{\nu}$	wegen des hohen Strahlungsdruckes ist Galaxienentstehung erst im durchsichtigen Kosmos möglich
	10^{10} a	Bildung des Sonnensystems; γ, ν, $\bar{\nu}$	
$T_{\gamma_0} = 2{,}7$	$1{,}8 \cdot 10^{10}$ a		heutiger Zustand

Restbeitrag zur Gesamtenergie, alle Energien auf $V = K^3$ bezogen). Damit ergibt sich aus $\dot{E} + p\dot{V} = 0$:
$\dot{E}_G$ = konst. $= Mc^2 = \varrho_0 c^2 K^3 = \varepsilon_0 K^3$, $\dot{E}_R + p_R \dot{V} = 0$, $\dot{E}_\gamma + p_\gamma \dot{V} = 0$. Die Adiabatengleichung der Hintergrundstrahlung liefert $(E_\gamma = \varepsilon_\gamma(T) K^3$, $p_\gamma = [1/3]\, \varepsilon_\gamma)$ als Entropie der 3-K-Strahlung $S_\gamma = [4/3]\, E_\gamma / T_\gamma = 4\, a\, (T_\gamma K)^3/3$ = konst. ($\varepsilon_\gamma = a T_\gamma^4$ = Stefan-Boltzmann-Gesetz). Während der Expansion fällt also die Temperatur der Strahlung im umgekehrten Verhältnis zum Abstand zweier beliebiger Punkte des Kosmos $(T_\gamma \sim K^{-1})$. Aus S_γ läßt sich die in Einheiten der Boltzmann-Konstante k_B ausgedrückte Entropie je Baryon (Baryonen – Nukleonen + Hyperonen) σ berechnen, wobei man annimmt, daß die Nukleonen den Hauptanteil der Massendichte ϱ_0 bilden ($\varrho_0 \approx \varrho_N \equiv N_N m_N K^{-3} = MK^{-3}$, N_N = Gesamtzahl der Nukleonen, m_N = Ruhmasse eines Nukleons):

$$\sigma \equiv S_\gamma / k_B N_N = 4\, a T_{\gamma_0}^3 m_N / 3\, k_B\, \varrho_0 = 2{,}4 \cdot 10^{-21}\, \text{g}/\varrho_0\, \text{cm}^3 \approx 2{,}4 \cdot 10^9 \; (\varrho_0 \approx 10^{-30}\, \text{g/cm}^3).$$

Die konstante Entropie je Baryon ist das Bindeglied zwischen den einzelnen Entwicklungsetappen des Universums. Ihr hoher Wert entspricht einer großen Photonenzahldichte n_{γ_0} gegenüber der Baryonenzahldichte n_{N_0} ($\sigma = 3{,}7\, n_{\gamma_0}/n_{N_0}$). Aus diesem Grund heißt dieser Kosmos auch heißes Modell des Universums. Da die Photonen zu früheren Zeiten sehr energiereich waren $(T_\gamma \sim K^{-1})$, verhinderten sie die Bildung von Atomkernen (erst recht von Atomen), sie wechselwirkten damals so stark mit den Elementarteilchen (Streuprozesse, Paarerzeugung), daß die gesamte Materie im thermischen Gleichgewicht war. Welche Elementarteilchensorten sich über Erzeugungs- und Vernichtungsprozesse mit der elektromagnetischen Strahlung im Gleichgewicht befanden, wird durch das Verhältnis von mittlerer Energie der Photonen zu Ruhenergie der Teilchen bestimmt ($m_i c^2 \lesssim k_B T(t)$, i bezeichnet Teilchensorten im Gleichgewicht). Teilchen, zu denen es Antiteilchen gibt, verschwinden völlig, wenn $T_\gamma(t)$ infolge der Expansion den durch die Ruhmasse der Teilchensorte vorgegebenen Schwellwert unterschreitet (vgl. Tab. 3.5.6–1).

3.5.6. Temperaturgeschichte des Universums

Mit $T_\gamma(t) \sim K^{-1}(t)$ und $K(t)$ als Lösung der Friedman-Gleichung kann man zu jedem Zeitpunkt die Temperatur des Universums angeben (Tab. 3.5.6–1).
Die Beschreibung der ersten hundertstel Sekunde hat stark hypothetische Züge, dazu reichen die derzeit bekannten Naturgesetze nicht aus (notwendig: einheitliche Theorie aller Wechselwirkungen). Möglich wäre folgender Verlauf (Tab. 3.5.6–2). Die heutige 3-K-Strahlung stammt aus einer Zeit, zu der der Kosmos für elektromagnetische Strahlung

Tab. 3.5.6–2 Hypothese über die erste hundertstel Sekunde

T in Kelvin	t in Sekunden	Bemerkungen
$> 10^{12}$	0,01 s	Bildung von Hadronenpaaren (Hadronen – Baryonen, Mesonen) Zeitalter der starken Wechselwirkungen
		Komponenten: γ, Leptonen, Quarks
$3 \cdot 10^{15}$		Symmetrie elektromagnetischer und schwacher Wechselwirkungen – elektroschwache Wechselwirkungen
	10^{-24} s	Teilchenerzeugung im Gravitationsfeld
10^{32}	10^{-43} s	Gravitation im Gleichgewicht mit den übrigen Wechselwirkungen

durchsichtig wurde. Sie ermöglicht also den Blick weit in die kosmische Vergangenheit zurück. Die aus der Strahlung abgeleitete Entropie je Baryon hat entscheidende Bedeutung für die gesamte Entwicklung des Weltalls. Aus dieser Zahl kann auch das kosmologische H-He-Verhältnis bestimmt werden, das mit der Beobachtung gut im Einklang steht. Das heiße Modell liefert zusätzlich eine Neutrino- ($T_{\nu_0} \approx 1{,}9$ K) und eine Gravitationshintergrundstrahlung ($T_{g_0} \approx 1$ K). Für diese beiden Strahlungen wurde der Kosmos schon wesentlich früher durchsichtig (vgl. Tab. 3.5.6–1 und –2). Deshalb würde ihre experimentelle Untersuchung weitere Klarheit über die Anwendbarkeit des Modells für die Frühphase des Universums (z. B. Isotropie) bringen. Bisher konnten diese Strahlungen nicht registriert werden.

3.5.7. Zur Entropie je Baryon

Die Ursache der sehr hohen Entropie je Baryon ist noch nicht geklärt. Möglicherweise ist sie das Relikt einer hier nicht diskutierten kosmischen Entwicklungsetappe jenseits des Urknalls. Wahrscheinlicher erscheint, daß das mit dem Urknall entstandene kosmologische Modell anfangs ($t < 10^{-35}$ s) kein Friedman-Kosmos war und durch dissipative (entropieerzeugende) Prozesse (Teilchenerzeugung durch Ausgleich von Raumkrümmung, Neutrinoviskosität) in seiner großräumigen Struktur (lokale Störungen blieben eventuell bestehen – Galaxienkeime) zu einem Friedman-Universum geglättet wurde. Dies lieferte eine noch größere spezifische Entropie ($\sigma \approx 10^{40}$). Der gemessene Wert ($\sigma = 10^9$) könnte in einer darauffolgenden Phase durch baryonenzahlverletzende Prozesse zustande gekommen sein. Moderne Elementarteilchentheorien, die die elektromagnetische, die schwache und die starke Wechselwirkung im Rahmen einer einheitlichen Theorie (nichtabelschen Eichtheorie) beschreiben, ermöglichen baryonenzahlverletzende Prozesse. Damit wäre auch geklärt, warum im Kosmos z. B. Elektronen und Protonen in merklicher Anzahl vorhanden sind, Positronen und Antiprotonen kosmologischen Ursprungs dagegen nicht beobachtet werden (Baryonenasymmetrie).

3.5.8. Die Galaxienbildung

Sie ist im Detail noch unverstanden. Eine Theorie geht von einem homogenen isotropen Universum aus, in dem aus lokalen zufälligen Schwankungen (Fluktuationen) der Massendichte unter dem Einfluß der Gravitation und der elektromagnetischen Strahlung zunächst Dichteschwankungen entstanden sind. Diese mußten bereits in der Anfangsphase des Kosmos bis zu einer Mindestgröße anwachsen, um nicht durch die kosmologische Expansion und gewisse entropieerzeugende Prozesse bis zur Entkopplung von Ruhmasse und elektromagnetischer Strahlung ($t \approx 3 \cdot 10^5$ Jahre) völlig ausgedämpft zu werden. Die Entkopplung ermöglichte den Massen, sich aufgrund der nun fehlenden Behinderung durch den Strahlungsdruck zu massereichen Objekten zusammenzuballen. Ob dieser Prozeß dann allerdings schnell genug ablaufen konnte, ist fraglich. Ein zweiter Erklärungsversuch bezweifelt die Entstehung der heutigen Galaxien ($M \approx 10^{11} M_\odot$) aus zufälligen Schwankungen. Es wird vermutet, daß Inhomogenitäten (die Galaxienkeime) als Anfangsinformation im Urknall enthalten waren, die kosmische Entwicklung also das Ergebnis sehr spezieller Anfangsbedingungen ist und die Ruhmassendichte nahe der kritischen Massendichte liegt (Weltmodell sehr ähnlich einem Einstein-de-Sitter-Modell). Ein solcher Kosmos bleibt während seiner Entwicklung stabil und hat eine für die Galaxienbildung günstige Expansionsgeschwindigkeit. Eine zu hohe Geschwindigkeit würde einer Massenkonzentration zu stark entgegen wirken, bei einer zu niedrigen hätte sich die Materie nicht genügend abkühlen können, die für die Klumpung notwendige Entkopplung von Strahlung und Ruhmasse wäre nicht erfolgt. Anisotropien der 3-K-Strahlung infolge kleinräumiger Störungen zum Zeitpunkt der Entkopplung liegen unter der derzeitigen Meßgenauigkeit ($3 \cdot 10^{-4}$). Die Störungen (Galaxienkeime) waren zu diesem Zeitpunkt also außerordentlich klein.
Trotz der noch ungeklärten Probleme (Entropie je Baryon, Baryonenasymmetrie, Grenzen des kosmologischen Prinzipes, Singularität des Anfangszustandes) gibt das heiße (Standard-)Modell des Universums Erklärungen für die 3-K-Strahlung, die kosmologische Rotverschiebung und das H-He-Verhältnis. Der für die zukünftige Entwicklung des Kosmos oft vorhergesagte Wärmetod (stabiler thermodynamischer Gleichgewichtszustand) ist im Rahmen der Einsteinschen Gravitationstheorie nicht möglich. Die Gravitation verhindert einen solchen Endzustand. Es existiert keine andere Gravitationstheorie, die Modelle zuläßt, die mit vergleichbarer Klarheit die kosmologischen Beobachtungen auf einheitlicher Basis deuten.

3.6. Allgemeine Feldtheorie

3.6.1. Begriffliche Einführung

In der nichtrelativistischen klassischen kanonischen Mechanik eines Systems von N Massenpunkten werden diese Massenpunkte durch die Lagekoordinaten $q_\Omega(t)$ und die Impulskoordinaten $p_\Omega(t)$ charakterisiert, wobei Ω die Anzahl der Freiheitsgrade ($\Omega = 1, \ldots, 3N$) durchläuft und t die Newtonsche absolute Zeit bedeutet, die die Rolle eines Parameters spielt. Die Lage- und Impulskoordinaten gehören dabei einer kommutativen Algebra an.
In der zugeordneten nichtrelativistischen Quantenmechanik werden diese Lage- und Impulskoordinaten zu Operatoren, die eine nichtkommutative Algebra bilden. Es gelten für sie die Heisenbergschen Vertauschungsregeln

a) $[q_\Omega, q_\Gamma] = 0$, b) $[p_\Omega, p_\Gamma] = 0$, c) $[q_\Omega, p_\Gamma] = i\hbar\delta_\Omega$ (Ω, $\Gamma = 1, \ldots, 3N$). Dabei bedeuten die Klammerausdrücke Kommutatoren. Diese Operatoren wirken auf Zustände des

Hilbert-Raumes, die man durch Ket-Vektoren $|\Psi\rangle$ bzw. Bra-Vektoren $\langle\Psi|$ symbolisiert.
Während es sich bei der klassischen und Quantenmechanik – im Unterschied zur Kontinuumsmechanik – um Systeme mit endlich vielen Freiheitsgraden handelt, liegen in der Feldtheorie Systeme mit nicht-abzählbar unendlich vielen Freiheitsgraden vor. Dabei wird ein Feldsystem, das i. allg. aus verschiedenen Feldarten besteht, die sich in der Regel in Wechselwirkung befinden, durch die N Feldfunktionen (Tensoren oder Spinoren) $U_\Omega = U_\Omega(x^i)$ beschrieben ($\Omega = 1, \ldots, N$), wobei die x^i die vierdimensionalen Raum-Zeit-Koordinaten ($i = 1, \ldots, 4$) sind. Zwischen Punktmechanik und Feldtheorie besteht die Zuordnungsanalogie
$q_\Omega(t) \rightarrow U_\Omega(x^i)$, $t \rightarrow x^i$. In der Speziellen Relativitätstheorie handelt es sich bei den x^i um die sogenannten Galilei-Koordinaten, die die räumlichen kartesischen Koordinaten $\{x^\alpha\} = (x, y, z)$ und die Inertialzeit t umfassen. In der Feldtheorie hängen also die Feldfunktionen U_Ω neben der Zeit auch noch von den kontinuierlichen räumlichen Koordinaten x^α ab, woraus sich die nichtabzählbar unendlich vielen Freiheitsgrade erklären.
In der klassischen Feldtheorie kommutieren die Feldfunktionen miteinander, bilden also eine kommutative Algebra. In der Quantenfeldtheorie gehören die Feldfunktionen als Operatoren einer nichtkommutativen Algebra an, wobei für sie spezifische Vertauschungsregeln gelten. Die Raum-Zeit-Koordinaten selbst bleiben auch in der Quantenfeldtheorie kommutative Parameter.
Man teilt die Feldtheorien in lineare Feldtheorien (lineare Feldgleichungen) und nichtlineare Feldtheorien (nichtlineare Feldgleichungen) ein.
Die strenge mathematische Beherrschung der Quantenfeldtheorie gekoppelter Felder ist heute noch aktueller Forschungsgegenstand, da gewisse charakteristische Schwierigkeiten bisher nicht völlig überwunden werden konnten (z.B. Divergenzen, Konvergenzschwierigkeiten usw.). Aber auch eine Reihe physikalischer Fragen sind noch nicht endgültig geklärt (z. B. Vertauschungsregeln, Definitheit der Metrik des Hilbert-Raumes usw.). Die weitere Erforschung und das bessere Verständnis der Quantenfeldtheorie bilden eine unabdingbare Voraussetzung für den Aufbau einer zukünftigen Theorie der Elementarteilchen, die mit großer Wahrscheinlichkeit eine nichtlineare Quantenfeldtheorie sein wird.
Die bis heute gut bewährte Praxis geht, da bei der mathematischen Behandlung gewisser nichtkommutativer Spezifika der Quantenfeldtheorie noch viele Fragen ungeklärt sind, oft so vor, daß die theoretischen Formulierungen im Rahmen der zugeordneten klassischen Feldtheorien abgewickelt werden und danach die formale Übersetzung in die Begriffe der Quantenfeldtheorie erfolgt (Quantisierungsprozedur).

3.6.2. Einige konkrete Feldtheorien

Bei den physikalisch gesicherten und im wesentlichen abgeschlossenen Feldtheorien handelt es sich um nachfolgende Theoriengebäude.
3.6.2.1. Maxwell-Gleichungen. Die primären Feldfunktionen des Elektromagnetismus sind das elektromagnetische Viererpotential $\boldsymbol{A}_i$ ($i = 1, \ldots, 4$), das sich aus dem Vektorpotential $\boldsymbol{A}$ und dem skalaren Potential φ aufbaut. Durch Differentiation lassen sich daraus die elektrische Feldstärke $\boldsymbol{E}$ und die magnetische Feldstärke (magnetische Induktion) $\boldsymbol{B}$ aufbauen. In einem elektromagnetischen Medium mit den Eigenschaften der Dielektrizität und Permeabilität sind diesen beiden Feldstärken die elektrische Erregung (dielektrische Verschiebung) $\boldsymbol{D}$ und die magnetische Erregung (magnetische Feldstärke) $\boldsymbol{H}$ zugeordnet. (Die in Klammern aufgeführten Begriffe entsprechen der älteren traditionellen Begriffsbildung). Die elektromagnetischen Feldgrößen sind Tensoren und genügen den Maxwell-Gleichungen (lineare partielle Differentialgleichungen 1. Ordnung), die die Feldgleichungen für das Gebiet des Elektromagnetismus (inklusive Optik) sind. Die Quantisierung des Maxwell-Feldes führt zu den Photonen als den dem Maxwell-Feld zugeordneten Quantenteilchen.
3.6.2.2. Schrödinger-Gleichung, Klein-Gordon-Gleichung und Dirac-Gleichung. Die primäre Feldfunktion des Schrödinger-Feldes ist die komplexe skalare Wellenfunktion Ψ, die die spezifische Bewegung von nichtrelativistischen Quantenteilchen beschreibt. Sie hat die physikalische Bedeutung eines Wahrscheinlichkeitsfeldes und genügt der nichtrelativistischen Schrödinger-Gleichung (Äquivalenz dieser Schrödingerschen Wellenmechanik mit der Heisenbergschen Matrizenmechanik). Die Schrödinger-Gleichung ist eine lineare partielle Differentialgleichung 2. Ordnung.
Die relativistische Verallgemeinerung der Wellenmechanik erfolgte in folgenden zwei verschiedenen Richtungen:
Die Klein-Gordon-Feldfunktion ist ebenfalls eine komplexe skalare Feldfunktion. Sie genügt der Klein-Gordon-Gleichung, die eine lineare partielle Differentialgleichung 2. Ordnung ist. Wegen Indefinitheitseigenschaften ist dieses Feld für die Wahrscheinlichkeitsinterpretation in der Quantenphysik nicht geeignet.
Dieser Forderung entspricht das Diracsche Bispinor-Feld, das der relativistischen Dirac-Gleichung genügt, die eine lineare partielle Differentialgleichung 1. Ordnung ist.
Die Quantisierung des Schrödinger-Feldes führt zu den nichtrelativistischen Quantenteilchen allgemein, während die Quanten des Klein-Gordon-Feldes die skalaren Mesonen (Spin 0) und die Quanten des Dirac-Feldes die Elektronen und Positronen (Spinteilchen mit Spin 1/2) sind.
Eine herausragende und experimentell ausgezeichnet bestätigte Quantenfeldtheorie ist die Quantenelektrodynamik als Wechselwirkungstheorie von Maxwell-Feld und Dirac-Feld (Wechselwirkung von Photonen, Elektronen und Positronen). In neuerer Zeit ist zur Erklärung von Elementarteilchen-Phänomenen in Analogie dazu die sogenannte Quantenchromodynamik ausgearbeitet worden, deren Erforschung aber noch im Fluß ist.
3.6.2.3. Newtonsche und Einsteinsche Gravitations-Feldgleichungen. Beim Newtonschen Gravitationsfeld handelt es sich um ein reelles skalares Feld (Gravitationspotential), das einer nichtrelativistischen linearen partiellen Differentialgleichung 2. Ordnung genügt (Poisson-Gleichung).
Die allgemein-relativistische Verallgemeinerung des Newtonschen Gravitationspotentials ist das Einsteinsche metrische Feld, das den nichtlinearen Einsteinschen Gravitations-Feldgleichungen (System partieller Differentialgleichungen 2. Ordnung) genügt (vgl. 3.3.).

3.6.3. Hamilton-Prinzip, Lagrange-Gleichungen und Hamilton-Formalismus

Ähnlich der kanonischen Mechanik lassen sich die physikalisch bewährten Feldgleichungen im Rahmen der soge-

nannten lokalen Feldtheorien als Lagrange-Gleichungen aus dem feldtheoretischen Hamilton-Prinzip

$$\delta \int_{V_4} \mathcal{L}(U_\Omega, U_{\Omega,i}, x^i)\, \mathrm{d}^{(4)}V = 0$$

ableiten. Dieses Variationsprinzip ist ein Extremalprinzip, bei dem die Wirkung für die real ablaufende Bewegung der Felder zu einem Extremum gemacht wird. Es bedeuten: $\mathcal{L}$ = Lagrange-Dichte, $\mathrm{d}^{(4)}V$ = vierdimensionales Volumelement, V_4 = festes vierdimensionales Raum-Zeit-Gebiet, $U_{\Omega,i}$ = partielle Ableitung der Felder U_Ω nach x^i.
Notwendige Bedingung für die Extremierung der Wirkung ist die Erfüllung der Lagrange-Gleichungen

$$\partial\mathcal{L}/\partial U_\Omega - (\partial/\partial x^i)(\partial\mathcal{L}/\partial U_{\Omega,i}) = 0,$$

die als partielle Differentialgleichungen gerade die Feldgleichungen für die Felder U_Ω sind. Da es sich bei dem Hamilton-Prinzip (s. o.) aus physikalischen Gründen (Erfüllung des Relativitätsprinzips) um ein relativistisch kovariantes Prinzip handeln muß, werden die Lagrange-Gleichungen als Konsequenz davon zu kovarianten (forminvarianten) Gleichungen.
Parallel zu dem eben skizzierten Lagrange-Formalismus läßt sich ein feldtheoretischer Hamilton-Formalismus mit feldtheoretischen kanonischen Hamilton-Gleichungen aufbauen. Den Feldfunktionen U_Ω (Analogie zu den Lagekoordinaten q_Ω) lassen sich dabei Impulsfunktionen Π^Ω (Analogie zu den Impulskoordinaten p^Ω) zuordnen. Bei dieser Prozedur wird aber die Zeit explizit ausgezeichnet, so daß der feldtheoretische Hamilton-Formalismus, wenn man von Kunstgriffen absieht, seinem Wesen nach nicht relativistisch ist. Dennoch hat er für die Quantisierung der Felder eine fundamentale Bedeutung, da für U_Ω und Π^Ω in Analogie zu den Heisenbergschen Vertauschungsregeln die gleichzeitigen Vertauschungsregeln gelten:

a) $[U_\Omega(x^\alpha, t), U_\Gamma(\bar{x}^\alpha, t)]_\pm = 0,$
b) $[\Pi^\Omega(x^\alpha, t), \Pi^\Gamma(\bar{x}^\alpha, t)]_\pm = 0,$
c) $[U_\Omega(x^\alpha, t), \Pi^\Gamma(\bar{x}^\alpha, t)]_\pm = \delta^\Gamma_\Omega\, \delta^{(3)}(x^\alpha - \bar{x}^\alpha)$

(Plus-Quantisierung für Fermionen, Minus-Quantisierung für Bosonen), wobei δ^Γ_Ω das Kronecker-Symbol und $\delta^{(3)}(x^\alpha - \bar{x}^\alpha)$ die dreidimensionale räumliche Diracsche Deltafunktion sind.

3.6.4. Symmetrie, Erhaltung und Noether-Theorem

Es besteht ein tiefgründiger physikalischer Zusammenhang zwischen den Symmetrien physikalischer Feldsysteme und den physikalischen Erhaltungsgesetzen. Grundsätzlich unterscheidet man zwischen der Klasse der diskreten Symmetrien und der Klasse der kontinuierlichen Symmetrien. Konkret bedeutet das, daß bei entsprechenden Symmetrietransformationen die Lagrange-Dichte (eventuell bis auf eine unwesentliche Divergenz) ihre Form behält.
Die wichtigsten diskreten Symmetrietransformationen sind:
- Raumspiegelung (führt zur Erhaltung der Raumparität),
- Zeitumkehr,
- Ladungskonjugation (führt zur Erhaltung der Ladungsparität).

Da es sich bei den Paritätsbegriffen um typische quantenphysikalische Phänomene handelt, die nur im Rahmen des Operatorkalküls mathematisch zugänglich sind, bezieht sich diese Klasse von Symmetrietransformationen primär auf die Quantenphysik. Die tiefe innere Verbindung dieser Symmetrien widerspiegelt das PTC-Theorem von W. Pauli und G. Lüders (etwa 1952).
Historisch eingeleitet wurden diese physikalischen Einsichten über den Zusammenhang von Symmetrie und Erhaltung durch das Studium der kontinuierlichen Symmetrien von H. Poincaré für die Mechanik und E. Noether (1918) für die Feldtheorie.
Die Klasse der kontinuierlichen Symmetrietransformationen teilt man in zwei grundsätzlich verschiedene Unterklassen ein:
- infinitesimale Koordinaten-Transformationen (äußere Symmetrietransformationen) mit ihren Auswirkungen auf die tensoriellen und spinoriellen Transformationsgesetze der Feldfunktionen U_Ω (Angreifen der Transformationen an den Tensor- oder Spinorindizes Ω),
- infinitesimale Funktionstransformationen als erweiterte Eichtransformationen (innere Symmetrietransformationen), bei denen die Koordinaten unverändert bleiben, aber die Struktur der Funktionen selbst infinitesimal abgeändert wird.

Die wichtigsten Zuordnungen von Symmetrie und Erhaltung im Rahmen der speziellen Relativitätstheorie sind:

- räumliche Translation → Impulserhaltung
- zeitliche Translation → Energieerhaltung

} Homogenität der Raum-Zeit (vier Erhaltungssätze),

- räumliche Drehung → Drehimpulserhaltung
- raumzeitliche „Drehung“ („Lorentz-Drehung“, Boost) → Schwerpunkterhaltung

} Isotropie der Raum-Zeit (sechs Erhaltungssätze),

Innere Symmetrien:
- elektromagnetische Eich-Phasen-Transformation → elektrische Ladungserhaltung (ein Erhaltungssatz).

4. Quantenphysik und Statistik

Quantenphysik ist die Physik des Mikroskopischen, die Physik im atomaren Bereich, während die uns im täglichen Leben begegnenden makroskopischen Systeme mit den Begriffen der klassischen Physik beschrieben werden können.
Die grundlegende Frage der Quantenphysik – Kontinuität oder Diskontinuität – hat schon die Philosophen des Altertums beschäftigt. Kann man einen Stoff beliebig fein unterteilen oder gelangt man zu einem kleinsten Element? Bezüglich der Masse und der Ladung ist uns heute schon geläufig, daß physikalische Größen gequantelt sind. Der kleinste Baustein eines chemischen Elementes ist ein Atom. Die Atommassen wurden experimentell ermittelt und liegen tabelliert vor. Die kleinste Ladungsmenge, die Elementarladung, ist die Ladungsmenge, die ein Elektron als negative Ladung bzw. ein Proton als positive Ladung trägt. Teile von Elektronen oder Teile der Elementarladung konnten nicht beobachtet werden.
Das Besondere der Quantenphysik besteht darin, daß die Diskontinuität auf andere physikalische Größen erweitert wird, z. B. auf Energie und Drehimpuls. Ein Atom hat diskrete Energiezustände und kann somit seine Energie nur sprunghaft, um endliche Energiequanten, ändern. Der Vektor Drehimpuls unterliegt einer Richtungsquantelung. Seine Projektion auf eine vorgegebene Richtung kann nur diskrete Werte annehmen. Die Ausnahmen bilden wohl nur Länge und Zeit. Es ist nicht bekannt, ob es eine Elementarlänge und eine Elementarzeit gibt. Die geringe Anschaulichkeit der Quantenmechanik liegt darin begründet, daß einige in der klassischen Mechanik gewohnte Begriffe, insbesondere der der Bahnkurve, nicht mehr verwendbar sind. Ein Elektron in einem Atom hat zwar eine bestimmte Energie und einen bestimmten Drehimpuls, aber es bewegt sich nicht auf einer Bahnkurve. Sein Zustand wird stattdessen durch eine Wellenfunktion charakterisiert. Die Unbestimmtheitsrelation bringt diesen Unterschied zur klassischen Physik zum Ausdruck. Sie stellt aber keine Einschränkung an die Erkennbarkeit der Natur dar. Es ist falsch zu sagen, die Bahnkurve sei nicht genau meßbar oder erkennbar. Die Eigenschaft Bahnkurve hat das Elektron im Atom nicht.
Das Bohr-Sommerfeldsche-Atommodell war deshalb so erfolgreich, weil die Bohr-Sommerfeldsche-Quantenbedingung Aussagen über die Erhaltungsgrößen Energie, Drehimpuls usw. macht, nicht weil es die richtigen „erlaubten Bahnen" auswählt. Die Existenz von Erhaltungsgrößen ist mit Invarianzeigenschaften des Systems verknüpft, die unabhängig davon existieren, ob man das System als klassisch oder nichtklassisch betrachten muß. Die im Bohr-Sommerfeldschem-Atommodell noch enthaltenen „erlaubten Bahnen" verschwinden dann auch bei der endgültigen Formulierung der Grundgleichung der Quantenmechanik durch Heisenberg und Schrödinger. Das Bohr-Sommerfeldsche-Atommodell wird bei einfachen Problemen gelegentlich der exakten quantenmechanischen Behandlung vorgezogen, weil es angeblich anschaulich ist. Das geht natürlich nur dort gut, wo das Modell gültig ist, d.h. beim Einelektronenproblem, wobei auch hier Postulate verwendet werden müssen, z. B. über die Stabilität gegen Abstrahlung oder die Auswahlregeln, deren Inhalt die exakte quantenmechanische Behandlung automatisch liefert. Zum anderen ist das Bohr-Sommerfeldsche-Atommodell nur scheinbar anschaulicher. Man ordnet dem Elektron im Atom Eigenschaften eines klassischen Systems zu, die es gar nicht aufweist. Dem Verständnis der Begriffsbildungen der Quantenmechanik ist ein solches Vorgehen auf jeden Fall abträglich. Daher wird dieses Modell auch nur kurz behandelt.
In diesem Teil werden die Begriffsbildungen der Quantenmechanik erläutert. An einfachen Beispielen wird demonstriert, daß die Schrödinger-Gleichung auf die für die Quantenphysik typischen Aussagen führt, auf die Existenz diskreter Energiezustände gebundener Teilchen. Das Paradebeispiel ist das Wasserstoffatom. Die genaue Kenntnis dieses Beispiels ist Voraussetzung für das Verständnis des Aufbaus der Atome. Die Diskussion der vielfältigen Anwendungen der Quantentheorie auf Atome, Moleküle und Festkörper bleibt anderen Teilen des Buches überlassen. Daher treten in diesem Teil auch die Aspekte der Theoretischen Physik stärker hervor.
Die statistische Physik entwickelte sich mit der Festigung der Vorstellung vom atomaren Aufbau der Materie. Ludwig Boltzmann war einer der Pioniere. Ein makroskopisches System enthält größenordnungsmäßig 10^{23} Teilchen (vgl. Loschmidtsche Zahl). Daher muß von der Mathematik her das Rechnen mit großen Zahlen übernommen werden. Die statistische Physik ist nicht einfach die Aufsummation der Bewegung von 10^{23} Teilchen. In Systemen mit großer Teilchenzahl kommen Gesetze zum Vorschein, die von der Bewegung eines einzelnen Teilchens her unbekannt sind. Erinnert sei nur an das Boltzmannsche H-Theorem und die damit verbundene Einführung der Entropie. Ausgehend von der Beschreibung des Systems mit vielen Freiheitsgraden besteht das Ziel der statistischen Physik darin, Aussagen über die meßbaren makroskopischen Größen zu erhalten, d.h. über die thermodynamischen Zustandsgrößen und Korrelationsfunktionen. Da die Statistik sich mit mikroskopischen Objekten befaßt, sind gegebenenfalls die Besonderheiten der Quantenmechanik zu beachten. Daher wird neben der klassischen Statistik auch die Quantenstatistik, d.h. die Fermi-Dirac- und Bose-Einstein-Statistik, behandelt.

4.1. Grundlagen der Quantenmechanik

Um die Jahrhundertwende mehrten sich experimentelle Ergebnisse (Fotoeffekt, Hohlraumstrahlung, diskrete Atomspektren), die aus den Grundgleichungen der klassischen Physik nicht verstanden werden konnten bzw. im Widerspruch zu ihnen standen. Besonders intensiv wurde die Hohlraumstrahlung untersucht. Die Herleitung der Formel für das spektrale Emissionsvermögen führte Planck zur Hypothese der *Energiequanten* und damit zur Entdeckung einer neuen Naturkonstanten, dem Planckschen Wirkungsquan-

tum h. Die Weiterentwicklung der Planckschen Idee durch EINSTEIN, BOHR und SOMMERFELD gipfelte in der Formulierung der Grundgleichungen der Quantenphysik durch HEISENBERG und SCHRÖDINGER.

4.1.1. Teilchen-Welle-Dualismus

Mit einer *Welle* verknüpft man die Vorstellung eines ausgedehnten Wellenfeldes, für das Beugungs- und Interferenzerscheinungen typisch sind. Diese Vorstellung verband man um die Jahrhundertwende mit dem Licht, besonders nach dem überwältigenden Erfolg der Maxwellschen Theorie elektromagnetischer Wellen. Unter einem Teilchen versteht man ein räumlich lokalisiertes Objekt, dessen Bewegung durch eine Bahnkurve charakterisiert wird. Von den Katodenstrahlen, d. h. den Elektronen, die man im elektrischen und magnetischen Feld ablenken konnte und für die man das e/m-Verhältnis bestimmen konnte, hatte man dieses Bild.

4.1.1.1. Fotoeffekt (Abb. 4.1.1–1). Hierbei beobachtete man, daß die Energieverteilung der aus einer Metalloberfläche ausgelösten Elektronen durch die Frequenz des einfallenden Lichtes, nicht durch dessen Intensität, bestimmt wird. EINSTEIN konnte dies 1905 mit der *Lichtquantenhypothese* – Licht der Frequenz ν besteht aus Quanten mit der Energie $E = h\nu = \hbar\omega (\hbar = h/2\pi)$ – erklären. Die Energie des Lichtquantes, das von einem Elektron absorbiert wird, teilt sich in Bindungsenergie W_B, Austrittsarbeit W_A und kinetische Energie auf $\hbar\omega = W_{kin} + W_A + W_B$. Mißt man die zu $W_B = 0$ gehörende maximale kinetische Energie über den Grenzwert U_{gr} einer Gegenspannung, dann gilt $\hbar\omega = eU_{gr} + W_A$. Für h ergaben die Messungen einen Wert, der mit dem bei der Hohlraumstrahlung gewonnenen übereinstimmte.

4.1.1.2. Compton-Effekt. Hierunter versteht man die von COMPTON Anfang der zwanziger Jahre beobachtete Erscheinung, daß bei der Beugung von Röntgen-Strahlen eine in den langwelligen Bereich verschobene Nebenlinie auftritt. Die Erklärung gelang unter der Annahme, daß *Energie und Impuls der Lichtquanten* durch Frequenz und Wellenlänge bestimmt sind,

$$E = \hbar\omega, \quad p = \hbar k$$

$(\omega = 2\pi\nu, k = 2\pi/\lambda)$, und ein elastischer Stoß zwischen Lichtquant und Elektron stattfindet, bei dem das Lichtquant einen Teil seiner Energie – je nach dem Streuwinkel ϑ – verliert. Energie- und Impulsbilanz führen in Übereinstimmung mit der beobachteten Wellenlängenänderung auf $\Delta\lambda = \lambda_c(1 - \cos\vartheta)$, mit der *Compton-Wellenlänge* $\lambda_c = h/mc$.

4.1.1.3. Elektronenbeugung, in quantitativ auswertbarer Form von DAVISSON und GERMER bzw. THOMSON 1927 gemessen, konnte mit der von DE BROGLIE entworfenen Vorstellung von *Materiewellen* verstanden werden. Danach werden die bei den Lichtquanten gefundenen Beziehungen benutzt, um einem Teilchen eine Welle mit einer aus $p = \hbar k$ folgenden Wellenlänge $\lambda = h/mv$, der *De-Broglie-Wellenlänge*, zuzuordnen. Hat ein Elektron seine kinetische Energie dadurch gewonnen, daß es ein Spannungsgefälle U durchlaufen hat, dann ergibt sich im nichtrelativistischen Bereich $\lambda = 12{,}25\,(U/\mathrm{V})^{-1/2}\,10^{-10}\,\mathrm{m}$ (Tab. 4.1.1–2). Von dieser Wellenlänge ausgehend erhielt man aus der Elektronenbeugung die gleichen Gitterparameter wie aus der Beugung von Röntgen-Strahlen.

Tab. 4.1.1–2 Wellenlänge von Elektronen verschiedener Energie

Energie in eV	Wellenlänge in nm	Energie in eV	Wellenlänge in nm
1	1,2264	200	$8{,}6713 \cdot 10^{-2}$
2	$8{,}6722 \cdot 10^{-1}$	500	$5{,}4834 \cdot 10^{-2}$
5	$5{,}4848 \cdot 10^{-1}$	1000	$3{,}8764 \cdot 10^{-2}$
10	$3{,}8783 \cdot 10^{-1}$	10^4	$1{,}2205 \cdot 10^{-2}$
20	$2{,}7424 \cdot 10^{-1}$	10^5	$3{,}7015 \cdot 10^{-3}$
50	$1{,}7344 \cdot 10^{-1}$	10^7	$1{,}1810 \cdot 10^{-4}$
100	$1{,}2264 \cdot 10^{-2}$	10^{10}	$1{,}2398 \cdot 10^{-7}$

4.1.1.4. Heisenbergsche Unbestimmtheitsrelation. Die Verknüpfung von Teilchen- und Welleneigenschaften führte zu der Vorstellung von Wellenpaketen (vgl. 2.3.3.), die sowohl eine mittlere Wellenlänge wie auch eine Lokalisierung aufweisen. Die von der Wellenlehre her bekannte Beziehung zwischen der Breite eines Wellenpaketes im

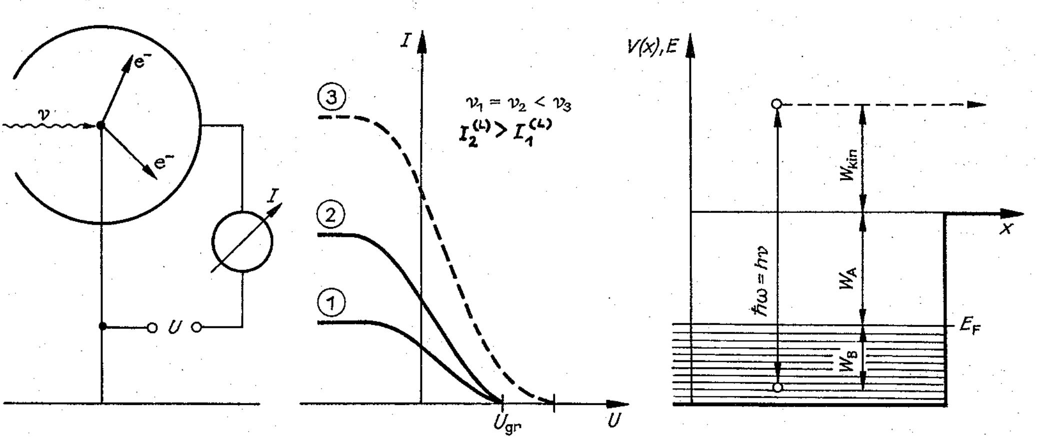


Abb. 4.1.1–1 Messung des Fotoelektronenstromes I in Abhängigkeit von der Gegenspannung U und Energiebilanz für das absorbierende Elektron

Ortsraum und im Wellenzahlraum führte mit der De-Broglie-Beziehung auf die *Heisenbergsche Unbestimmtheitsrelation*

$\Delta x \Delta p \geqq \hbar/2$, $h = 6{,}626 \cdot 10^{-34}\,\mathrm{Ws}^2$, $\hbar = 1{,}0546 \cdot 10^{-34}\,\mathrm{Ws}^2$.

Danach sind Ort und Impuls eines Teilchens prinzipiell nur mit einer begrenzten Genauigkeit gleichzeitig meßbar. Damit verliert die Bahnkurve eines Teilchens ihren in der klassischen Mechanik geprägten Sinn. Ein quantenmechanisches Teilchen hat diese Eigenschaft nicht. Die Unschärferelation ist also keine Einschränkung für die Erkennbarkeit der Natur. Behauptungen in dieser Richtung basieren auf dem Fehler, daß man einem System Eigenschaften zuordnen will, die es nicht hat.

Für makroskopische Objekte ergeben weder De-Broglie-Beziehung noch Unbestimmtheitsrelation neue oder einschränkende Gesichtspunkte. Mit $m = 1$ kg und $v = 1$ m/s ergibt sich $\lambda = 6{,}6 \cdot 10^{-34}$ m. Beugungserscheinungen, d. h. Welleneigenschaften, sind bei diesen kurzen Wellenlängen nicht beobachtbar. Weiterhin liefert die durch die vorhandenen Meßinstrumente gegebene Meßgenauigkeit von Ort und Geschwindigkeit ein Produkt $\Delta x \Delta p$, daß viele Zehnerpotenzen über $\hbar$ liegt. Die Unschärferelation liefert keine Einschränkung für die Messung von Ort und Impuls an Objekten der klassischen Mechanik. Anders liegen die Größenordnungen bei der Betrachtung eines Elektrons im atomaren Bereich. Wegen der kleinen Elektronenmasse und der geringen Ausdehnung eines Atoms stellt die Unbestimmtheitsrelation eine wesentliche Einschränkung der klassischen Mechanik dar. Im atomaren Bereich kann die Newtonsche Mechanik nicht verwendet werden, dort ist die Erweiterung zur Quantenmechanik erforderlich.

4.1.1.5. Bohr-Sommerfeldsche Quantenbedingung. Aufbauend auf die Plancksche Quantelung der harmonischen Bewegung und auf das Korrespondenzprinzip fanden Bohr und Sommerfeld eine halbklassische Beschreibung der Bewegung im atomaren Bereich, $\Delta I = h$, $I = \oint \mathrm{d}q\, p\,(q, E)$. Die Quantenbedingung für das Integral in der Phasenebene kanonisch konjugierter Variablen p, q sortiert aus den kontinuierlichen Werten der Erhaltungsgrößen der klassischen Mechanik diskrete durch die Forderung aus, daß die Phasenfläche zwischen zwei erlaubten Werten die Größe h hat. Sie führte für das Wasserstoffatom auf das beobachtete Termschema und wurde bis Ende der zwanziger Jahre erfolgreich zur Erklärung von Atomspektren benutzt. Sie ist auch heute für abschätzende Betrachtungen eine Hilfe.

4.1.2. Quantelung im Heisenberg-Bild

Die in der Unbestimmtheitsrelation enthaltene Aussage über die Einschränkung des Teilchenbildes konnte Heisenberg in eine allgemeingültige Forderung für kanonisch konjugierte Variable umformen. Führt man die *Minusvertauschung* $[p, q] = pq - qp$ ein, dann lautet die Forderung: Die Vertauschung kanonisch konjugierter Variablen hat den Wert $\hbar/i$. Die Bewegungsgleichungen der Quantenmechanik ergeben sich, indem man zu den Bewegungsgleichungen der klassischen Mechanik, den kanonischen Gleichungen, die Vertauschungsrelation hinzufügt. Bei einem Freiheitsgrad hat man

$\dot{p} = -\partial H/\partial q$, $\dot{q} = \partial H/\partial p$, $[p, q] = \hbar/i$.

Die *Vertauschungsrelation* führt zu einer wesentlichen Modifikation. Bei Zahlen ist die Reihenfolge im Produkt gleichgültig, sie sind vertauschbar. In der Quantenmechanik müssen *physikalische Größen durch Operatoren* dargestellt werden, man kann z. B. an Matrizen denken, bei denen die Reihenfolge im Produkt wesentlich ist. Die Meßwerte sollen durch die *Eigenwerte* der Matrix und die Zustände durch die *Eigenvektoren* bestimmt sein. Mit den Eigenwerten erhält man eine diskrete Folge von Meßwerten und damit eine Quantelung. Da Meßwerte reelle Größen sind, können physikalische Größen nur durch hermitesche Operatoren (vgl. 13.1.5.4.) beschrieben werden.

Die Bewegungsgleichungen lassen sich in die äquivalente Form

$\dot{A} = (i/\hbar)\,[H, A]$, $\quad A^{+} = A$, $\quad [p, q] = \hbar/i$

bringen, worin $A(p, q)$ eine beliebige physikalische Größe beschreibt. Der Hamilton-Operator – er geht aus der Hamilton-Funktion $H(p, q)$ hervor, indem man für p und q die Operatoren einsetzt – bestimmt den Zeitablauf physikalischer Größen. Hat der Hamilton-Operator für ein Teilchen die Eigenwerte – die Energieniveaus – E_k, dann ist für ein *System wechselwirkungsfreier Teilchen* die übersichtlichste Form des Hamilton-Operators

$H = \sum_k E_k a_k^{+} a_k + E_0$, $\quad [a_k, a_{k'}^{+}] = \delta_{k,k'}$.

k ist eine Quantenzahl, mit der die verschiedenen Zustände, in denen sich ein Teilchen befinden kann, durchnummeriert werden. $a_k^{+} a_k$ ist ein Operator mit den Eigenwerten $0, 1, 2, \ldots$, $\sum_k a_k^{+} a_k$ wird als *Teilchenzahloperator* bezeichnet. In einem Zustand eines Mehrteilchensystems mit dem Eigenwert n_k von $a_k^{+} a_k$ befinden sich n_k Teilchen in dem Einteilchenzustand k mit der Energie E_k. Die Eigenwerte des Hamilton-Operators $E = \sum_k E_k n_k$ lassen sich anschaulich verstehen. Die Energie E_k eines Niveaus wird mit der Zahl der Teilchen in diesem Niveau multipliziert, dann wird die Summe über alle Niveaus geführt. Wirkt der Operator a_k^{+} auf eine Eigenfunktion mit n_k Teilchen im Zustand k, dann entsteht eine Eigenfunktion mit $n_k + 1$ Teilchen im Zustand k. Der Operator a_k führt dagegen auf einen Zustand mit $n_k - 1$ Teilchen. Daher werden a_k^{+} bzw. a_k *Erzeugungs-* bzw. *Vernichtungsoperatoren* genannt. a_k^{+} erzeugt ein Teilchen im Zustand k, a vernichtet ein Teilchen im Zustand k.

4.1.3. Schrödinger-Gleichung

Freie Felder werden durch eine Wellengleichung (vgl. 2.3.2.) beschrieben, die auf eine Dispersionsbeziehung $\omega = \omega(\boldsymbol{k})$ führt. Freie Teilchen werden durch eine *Energie-Impuls-Beziehung* $E = H(\boldsymbol{p})$ charakterisiert, die durch die Hamilton-Funktion $H(\boldsymbol{p}) = c\sqrt{(mc)^2 + \boldsymbol{p}^2} = mc^2 + \boldsymbol{p}^2/2m + \ldots$ gegeben ist. Sollen Teilchen- und Wellenbeschreibung gleichermaßen zutreffend sein und wird die Brücke durch die *Einstein-* und *De-Broglie-Beziehung* $E = \hbar\omega$ bzw. $\boldsymbol{p} = \hbar\boldsymbol{k}$ hergestellt, dann muß $\omega = \omega(\boldsymbol{k})$ in $E = H(\boldsymbol{p})$ übergehen, wenn kein Widerspruch auftreten soll. Für Photonen führt die Wellengleichung $\Box U = 0$ auf $\omega = c\,|\boldsymbol{k}|$. Daraus folgt $E = c\,|\boldsymbol{p}|$, was mit der relativistischen Energie-Impuls-Beziehung übereinstimmt, wenn man Teilchen der Ruhmasse $m = 0$ betrachtet.

Versucht man eine Wellengleichung zu konstruieren, die auf $E = \boldsymbol{p}^2/2m$ führt, dann gelangt man zur Schrödinger-Gleichung $(-\hbar/i)\,\partial/\partial t\,\Psi(\boldsymbol{r}, t) = (2m)^{-1}(\hbar/i)^2 \Delta\Psi(\boldsymbol{r}, t)$, da für ebene Wellen die Differentiationen $(-\hbar/i)\,\partial/\partial t \rightarrow E$ bzw. $(\hbar/i)\,\partial/\partial \boldsymbol{r} \rightarrow \boldsymbol{p}$ Faktoren Energie bzw. Impuls liefern. Die

Erweiterung auf die Bewegung im nichtkonstanten Potential lautet

$$-\frac{\hbar}{i}\frac{\partial}{\partial t}\Psi(r,t)=H(\check{p},r)\Psi(r,t),\quad \int d\tau\Psi^*\Psi=1,$$

$$H(\check{p},r)=\frac{\check{p}^2}{2m}+V(r),\quad \check{p}=\frac{\hbar}{i}\frac{\partial}{\partial r}.$$

$H(\check{p}, r)$ ist der *Hamilton-Operator*, den man aus der Hamilton-Funktion gewinnt, indem man den kanonisch konjugierten Impuls p durch den *Impulsoperator* $p = (\hbar/i)\,\partial/\partial r$ ersetzt.

Die *Normierungsbedingung* $\int d\tau\Psi^*\Psi = 1$ ergibt sich aus der *Bedeutung der Wellenfunktion*. Da in der Schrödinger-Gleichung die imaginäre Einheit i explizit auftritt, erhält man komplexe Wellenfunktionen als Lösung. Meßgrößen sind reell, daher kann die Wellenfunktion selbst keine unmittelbare physikalische Bedeutung haben. Die Diskussion führte zur *Wahrscheinlichkeitsinterpretation*. Die Wellenfunktion charakterisiert ein Ensemble gleichartiger Teilchen, sie macht Aussagen über den Mittelwert gleichartiger Messungen. Der reelle Ausdruck $|\Psi(r, t)|^2 d\tau$ gibt die Wahrscheinlichkeit an, zur Zeit t ein Teilchen im Volumenelement $d\tau$ in der Umgebung des Ortes r anzutreffen. Die Wahrscheinlichkeit, ein Teilchen irgendwo im Raum anzutreffen, ist Sicherheit, also Eins. Daraus resultiert die Normierungsbedingung $\int d\tau|\Psi(r, t)|^2 = 1$.

Die Wahrscheinlichkeitsinterpretation hat sich bewährt, alle experimentellen Ergebnisse lassen sich in diesem Bild verstehen. Man hat z. B. Elektronenbeugungsversuche mit einem Elektronenstrahl geringer Stromdichte durchgeführt. Das nächste Elektron wurde erst emittiert, wenn das vorangegangene schon den Auffangschirm erreicht hatte. Jedes Elektron flog also unabhängig von den anderen. Jedes Elektron schwärzte die Fotoplatte an einer lokalen Stelle. Nach geeigneter Dauer des Versuches ergab die Gesamtheit aller Schwärzungspunkte das aus der Wellenfunktion folgende Beugungsbild. Der Versuch zeigt auch, daß die Beschreibung des Elektrons mit einer ausgedehnten Wellenfunktion keinesfalls bedeutet, daß das Elektron selbst ausgedehnt ist. Sonst könnten Teile des Elektrons beobachtet werden. Schrödinger- und Heisenberg-Bild sind zwei äquivalente Beschreibungsformen quantenmechanischer Prozesse. Die *Matrixelemente* A_{mn} der Heisenberg-Operatoren ergeben sich aus dem Schrödinger-Bild zu $A_{mn} = \int d\tau\Psi_m^* A\Psi_n$. Der Impulsoperator $\check{p}_x = (\hbar/i)\,\partial/\partial x$ erfüllt die Vertauschungsrelation $[\check{p}_x, x] = \hbar/i$.

4.1.4. Physikalische Größen

In der klassischen Mechanik sind physikalische Größen Funktionen von Impuls, Ort und Zeit $A(p, r, t)$. In der Quantenmechanik treten an deren Stelle *hermitesche Operatoren*

$$A = A(\check{p}, r, t),\quad A = A^+,\quad \check{p} = (\hbar/i)\partial/\partial r.$$

Der hermitesch konjugierte Operator A^+ zu A ist über $\int d\tau\varphi^* A\Psi = \int d\tau(A^+\varphi)^*\Psi$ definiert. Gilt $A^+ = A$, dann nennt man den Operator hermitesch.

4.1.4.1. Erwartungswert einer physikalischen Größe

$$\bar{A}(t) = \int d\tau\Psi^*(r,t)A(\check{p},r)\Psi(r,t)$$

ist der mittlere Meßwert im Zustand $\Psi(r, t)$. Einzelmessungen schwanken um diesen Mittelwert. Die *mittlere quadratische Schwankung*

$$(\Delta A)^2 = \int d\tau\Psi^*(r,t)\,(A-\bar{A})^2\Psi(r,t) = \overline{A^2} - \bar{A}^2$$

charakterisiert die Streuung um den Mittelwert.

4.1.4.2. Eigenzustand $\chi(r)$ eines Operators $A(\check{p}, r)$ – oder der physikalischen Größe A – ist der Zustand, in dem die mittlere quadratische Schwankung verschwindet, $\Delta A = 0$. Man mißt also stets den gleichen Wert, den sogenannten Eigenwert. Die Eigenzustände sind Lösungen der Eigenwertgleichung

$$A(\check{p}, r)\chi_n(r) = a_n\chi_n(r),\quad \int d\tau\chi_n^*\chi_n = 1.$$

Dabei können nur *normierbare Lösungen* physikalische Zustände beschreiben. Hermitesche Operatoren, d. h. physikalische Größen, haben nur reelle Eigenwerte. Ihre Eigenfunktionen sind orthogonal zueinander, und sie bilden ein vollständiges Funktionssystem

$$\int d\tau\chi_n^*\chi_m = \delta_{nm},\quad \sum_n \chi_n(r)\chi_n^*(r') = \delta(r-r').$$

Für *vertauschbare Operatoren* $[A, B] = 0$ läßt sich ein *gemeinsames System von Eigenfunktionen* finden.

4.1.4.3. Zeitablauf von Erwartungswerten. Die Schrödinger-Gleichung führt auf

$$\dot{\bar{A}} = \int d\tau\Psi^*(i/\hbar)\,[H, A]\Psi + \int d\tau\Psi^*\partial A/\partial t\,\Psi.$$

Hängt der Erwartungswert nicht von der Zeit ab $\dot{\bar{A}} = 0$, dann nennt man die physikalische Größe eine Erhaltungsgröße. Hängt A nicht explizit von der Zeit ab $\partial A/\partial t = 0$, dann ist dies der Fall, wenn $[H, A] = 0$ gilt, wenn die physikalische Größe mit dem Hamilton-Operator *vertauschbar* ist. $A = 1$ charakterisiert die *Normierung*. Da eine Zahl mit H vertauschbar ist, ist die Normierung eine Erhaltungsgröße. Zu allen Zeiten ist die Wahrscheinlichkeit, ein Teilchen irgendwo im Raum anzutreffen, gleich Eins. Nur die Wahrscheinlichkeitsverteilung kann sich im Zeitablauf ändern. Da H mit sich selbst vertauschbar ist, ist die *Energie* eine Erhaltungsgröße, wenn H nicht explizit von der Zeit abhängt. Dies führt auf die Vereinfachung, daß die Schrödinger-Gleichung *stationäre Lösungen* hat, deren Zeitabhängigkeit

$$\Psi_n(r,t) = e^{-(i/\hbar)E_n t}\varphi_n(r),\ H\varphi_n = E_n\varphi_n,\ \int d\tau\varphi_n^*\varphi_n = 1$$

bekannt ist und deren Ortsanteil aus der sogenannten *zeitunabhängigen Schrödinger-Gleichung* zu bestimmen ist. In stationären Zuständen hängt die Wahrscheinlichkeitsdichte $|\Psi_n(r, t)|^2$ nicht von der Zeit ab. Die Lösung der zeitunabhängigen Schrödinger-Gleichung führt auf einen Satz von orthogonalen Eigenfunktionen φ_n mit zugehörigen Eigenwerten E_n, den Energieniveaus des Systems. Ausgehend von einem beliebigen *Anfangszustand* $\Psi_0(r)$ läßt sich der Zeitablauf $\Psi(r, t) = \sum_n \varphi_n(r)\int d\tau'\varphi_n^*(r')\Psi_0(r')e^{-(i/\hbar)E_n t}$ mittels der stationären Lösungen beschreiben.

4.1.4.4. Ehrenfestsche Sätze über den Zeitablauf der Schwerpunktsbewegung $\bar{r}(t) = \int d\tau\Psi^*(r,t)\,r\Psi(r,t)$ ergeben sich für den Hamilton-Operator $H = \check{p}^2/2m + V(r)$ unter Verwendung von $(i/\hbar)\,[H, r] = \check{p}/m$ und $(i/\hbar)\,[H, p] = -\partial V/\partial r = F(r)$. Der Schwerpunktimpuls ist durch den Erwartungswert des Impulses gegeben, $m\dot{\bar{r}} = \bar{p}$. Die Schwerpunktbeschleunigung wird durch den Erwartungswert der Kraft bestimmt, $m\ddot{\bar{r}} = \overline{F(r)} \approx F(\bar{r})$. Ist das Kraftfeld über die Ausdehnung des Wellenpaketes konstant, dann kann

der Erwartungswert der Kraft durch den Wert der Kraft am Schwerpunkt ersetzt werden. Der Schwerpunkt befolgt dann genau die Newtonsche Bewegungsgleichung, auch wenn das Wellenpaket im Zeitablauf auseinanderfließt (vgl. 2.3.3.).

4.1.5. Quantelung der Energie

Als wesentliche Einschränkung zur klassischen Mechanik ergab sich im atomaren Bereich das Auftreten diskreter Energiezustände. Daß die Schrödinger-Gleichung für gebundene Zustände zu einem diskreten Spektrum von Eigenwerten führt, soll jetzt ausführlich diskutiert werden.

4.1.5.1. Bewegung in einem Potentialtopf (Abb. 4.1.5–1). Für sie macht die zeitunabhängige Schrödinger-Gleichung $d^2\varphi(x)/dx^2 + k^2(x)\varphi(x) = 0$, $k^2 = (2m/\hbar^2)[E - V(x)]$ die Aussage, daß die Krümmung der Wellenfunktion im klassischen Aufenthaltsbereich $E \geqq V(x)$ negativ ist, dort die Wellenfunktion also einen cosinusartigen Verlauf hat. Außerhalb des klassischen Aufenthaltsbereiches $E \leqq V(x)$ ist die Krümmung positiv, die Wellenfunktion verläuft also exponentiell. Damit die Wellenfunktion einen physikalischen Zustand beschreiben kann, muß sie normierbar sein. Das ist nur dann der Fall, wenn sie beiderseits des klassischen Aufenthaltsbereiches exponentiell abfällt. Eine exponentiell anwachsende Funktion führt zu einem unendlich großen Wert des Normierungsintegrals, ist also nicht normierbar. Ist E_0 eine Energie, für die die Wellenfunktion diese Bedingung erfüllt (Abb. 4.1.5–1a), dann führen Werte $E = E_0 \pm \delta E$ zu nicht normierbaren Lösungen (Abb. 4.1.5–1b). Die Energie muß erst auf E_1 erhöht werden, bis wieder eine normierbare Lösung entsteht. Die Differentialgleichung hat für jedes E eine Lösung. Die Normierungsbedingung siebt diskrekte Energiewerte heraus, für die die Lösung normierbar ist, einen physikalischen Zustand beschreibt.

Aus der Lösung für den rechteckigen Potentialtopf sind folgende allgemeine Aussagen erkennbar: Die Energie des Grundzustandes ist nicht gleich dem Potentialminimum, es gibt eine *Nullpunktsenergie*. Die Wellenfunktion des Grundzustandes ist knotenfrei (kein Nulldurchgang). Bei jedem höher angeregten Niveau kommt ein *Knoten* hinzu. In einem symmetrischen Potential sind die Eigenfunktionen abwechselnd symmetrisch und antisymmetrisch.

4.1.5.2. Potentialkasten. Für $V(x) = 0$ für $|x| < a$, $V(x) = \infty$ für $|x| > a$ ergeben sich die Eigenwerte $E_n = (\hbar^2/2m)(n\pi/2a)^2$ mit $n = 1, 2, 3, \ldots$ Die Eigenfunktionen sind Cosinus- und Sinusfunktionen.

4.1.5.3. Linearer harmonischer Oszillator. Aus den in der Schrödinger-Gleichung $[-(\hbar^2/2m)d^2/dx^2 + m\omega^2x^2/2]\varphi(x) = E\varphi(x)$ auftretenden Konstanten lassen sich die Kombinationen $\hbar\omega$ bzw. $\hbar/m\omega = \beta^2$ mit den Dimensionen Energie bzw. (Länge)^2 bilden. Die Eigenwerte $E_n = \hbar\omega(n + 1/2)$ sind *äquidistant*. Diese besonders einfache Form des Eigenwertspektrums erleichterte PLANCK die Einführung seiner Quantenhypothese bei der Diskussion der Hohlraumstrahlung.

V(x), E [a.u.]
x [a.u.]
V(x)
E_5=−1,4913
E_4=−3,9415
E_3=−6,0719
E_2=−7,7730
E_1=−9,0054
E_0=−9,7507
$\varphi_4(x)$
φ_4
φ_0
φ_1
a

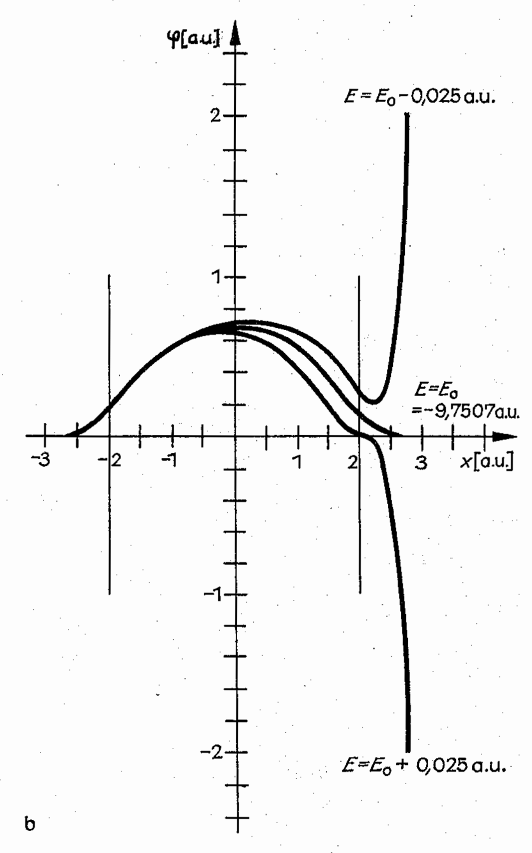


Abb. 4.1.5–1 (a) Eigenwerte und Eigenfunktionen im rechteckigen Potentialtopf, (b) nichtnormierbare Lösungen für $E = E_0 \pm \delta E$

Die Eigenfunktionen des harmonischen Oszillators $\varphi_n(x) = a_n H_n(x/\beta)\, e^{-(x/\beta)^2/2}$, $a_n^{-2} = 2^n n!\sqrt{\pi\beta}$ enthalten die hermiteschen Polynome (Abb. 4.1.5–2).

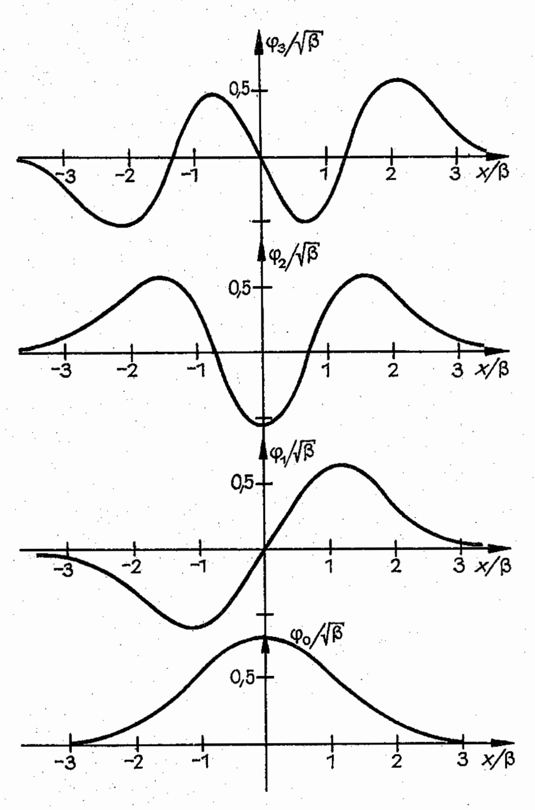


Abb. 4.1.5–2 Eigenfunktionen des harmonischen Oszillators

4.1.6. Zustände im Hilbert-Raum

Die in der Quantenmechanik betrachteten Zustände $\varphi(\boldsymbol{r})$ erfüllen die Bedingungen für die Elemente eines Hilbert-Raumes (vgl. 13.1.4.1.). Das ermöglicht es, viele Rechenoperationen durch analoge Betrachtungen im dreidimensionalen Vektorraum zu veranschaulichen. Im Unterschied zu diesem ist der Hilbert-Raum ein Vektorraum mit (i. allg.) unendlich vielen Dimensionen und mit komplexen Elementen. Jedes System von Eigenfunktionen χ_n eines hermiteschen Operators kann als orthogonale Basis des Hilbert-Raumes betrachtet werden. Die Entwicklung eines Zustandes φ nach dieser Basis $\varphi = \sum_n \chi_n c_n$ kann als Komponentenzerlegung des Hilbert-Vektors $|\varphi\rangle$ nach den Einheitsvektoren $|\chi_n\rangle$ aufgefaßt werden. Die Koeffizienten $c_n = \int d\tau \chi_n^* \varphi \equiv \langle\chi_n|\varphi\rangle$ ergeben sich aus dem Skalarprodukt $\langle\chi_n|\varphi\rangle$ zwischen $|\varphi\rangle$ und dem Einheitsvektor. Den Vektor mit den Komponenten c_n nennt man die Darstellung des Zustandes $|\varphi\rangle$ bezüglich der Basis $|\chi_n\rangle$. Wählt man als Basis die Eigenzustände des Hamilton-Operators, dann spricht man von der *Energiedarstellung*.

Lineare Operatoren ordnen einem Zustand einen anderen zu, $A\chi = \varphi$. Bei der Darstellung im Hilbert-Raum werden lineare Operatoren durch Matrizen mit den Elementen $A_{mn} \equiv \langle\chi_m|A|\chi_n\rangle = \int d\tau \chi_m^* A \chi_n$ beschrieben. Ein *unitärer Operator* $U^* = U^{-1}$ bewirkt eine Drehung der Basis. Mit ihm kann eine orthogonale Basis in eine andere orthogonale Basis überführt werden. $e^{-(i/\hbar)Ht}$ ist ein unitärer Operator. Die aus der Schrödinger-Gleichung folgende Lösung $|\Psi(t)\rangle = e^{-(i/\hbar)Ht}|\Psi(0)\rangle$ zeigt, daß im Hilbert-Raum der Zeitablauf des Zustandes als Drehung des Vektors $|\Psi(0)\rangle$ dargestellt wird.

Der Zeitablauf des Erwartungswertes $\bar{A}(t)$ eines Operators $A(\check{p}, \boldsymbol{r})$ ergibt sich im *Schrödinger-Bild* dadurch, daß die Zustände sich im Zeitablauf ändern $\bar{A}(t) = \langle\Psi(t)|A|\Psi(t)\rangle$. Im *Heisenberg-Bild* werden die Zustände als zeitunabhängig betrachtet. Dafür sind die Operatoren zeitabhängig $A(t) = e^{(i/\hbar)Ht} A(0)\, e^{-(i/\hbar)Ht}$, so daß sich die gleichen Erwartungswerte wie im Schrödinger-Bild ergeben.

4.2. Bewegung im Zentralkraftfeld

Die Bewegung im Zentralkraftfeld nimmt wie in der klassischen Mechanik auch in der Quantenmechanik wegen ihres häufigen Auftretens eine zentrale Stellung ein. Wie in der klassischen Mechanik ist es dabei möglich, unter Ausnutzung der Erhaltungsgrößen Energie und Drehimpuls die Aufgabe auf ein eindimensionales Bewegungsproblem zu reduzieren.

4.2.1. Erhaltungsgrößen

Im konservativen Zentralkraftfeld mit dem Potential $V(r)$ ist die *Energie eine Erhaltungsgröße*, da der Hamilton-Operator $H = \check{p}^2/2m + V(r)$ zeitunabhängig ist. Demzufolge gibt es *stationäre Lösungen* der Schrödinger-Gleichung mit bekannter Zeitabhängigkeit. Formt man den Hamilton-Operator in die Gestalt $H = p_r^2/2m + L^2/2mr^2 + V(r)$, $p_r = (\hbar/i)\,\partial/\partial r$, um und beachtet, daß der Drehimpulsoperator $\boldsymbol{L} = \boldsymbol{r} \times \boldsymbol{p}$ nur auf die Winkelkoordinaten operiert, dann erkennt man, daß $[H, \boldsymbol{L}] = 0$ gilt, daß der *Drehimpuls eine Erhaltungsgröße* ist. Verschiedene Drehimpulskomponenten sind nicht miteinander vertauschbar (vgl. 4.3.), aber L^2 und L_z. Wegen $[H, L^2] = 0$, $[H, L_z] = 0$ und $[L^2, L_z] = 0$ läßt sich ein gemeinsames System von Eigenfunktionen für die drei Operatoren H, L^2 und L_z finden. Die Drehimpulsoperatoren bestimmen dabei den *Winkelanteil der Wellenfunktion* $\varphi(\boldsymbol{r}) = (u_{n,l}/r)\, Y_{lm}(\vartheta, \lambda)$,

$$L^2 Y_{lm} = \hbar^2 l(l+1) Y_{lm}, \quad l = 0, 1, 2, \ldots;$$

$$L_z Y_{lm} = \hbar m Y_{lm}, \quad -l \leq m \leq l.$$

Die konkrete Form des Potentials $V(r)$ beeinflußt nur den Radialteil u/r der Wellenfunktion. $u(r)$ ist aus der *radialen Schrödinger-Gleichung* zu berechnen,

$$\left[-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + \frac{\hbar^2 l(l+1)}{2mr^2} + V(r)\right] u = Eu,$$

$$u(0) = 0, \quad \int_0^\infty dr\,|u|^2 = 1.$$

Für $r \to 0$ verläuft die normierbare Lösung $u \sim r^{l+1}$. $u(r)$ ist die Lösung einer Schrödinger-Gleichung für eine *eindi-*

mensionale Bewegung im Effektivpotential

$V_{eff}(r) = [\hbar^2 l(l+1)]/2mr^2 + V(r)$.

4.2.2. Wasserstoffatom

BALMER konnte 1885 zeigen, daß die Wellenlängen der vier im sichtbaren Bereich liegenden (damals bekannten) Spektrallinien des Wasserstoffes aus der Formel $1/\lambda = R_H(1/2^2 - 1/m^2)$, $m = 3, 4, 5, 6$, berechenbar sind (die Formel von BALMER wurde hier in eine für unsere Diskussionen zweckmäßige Form gebracht). RITZ faßte 1908 die bei der Auswertung von Spektren gefundenen Erfahrungen im sogenannten *Ritzschen Kombinationsprinzip* zusammen: Die reziproken Wellenlängen ergeben sich aus den Differenzen einer für jedes Atom typischen Termfolge $1/\lambda = T_n - T_m$, $m > n$. Danach gehört die Balmer-Serie zu $n = 2$. 1908 fand PASCHEN, dadurch angeregt, die zu $n = 3$ gehörende Serie im Ultraroten. Im *Bohrschen Atommodell* wurden diese Terme als Energieniveaus des Atoms interpretiert. Die von BOHR berechneten Energiewerte $E_n = -hcR_H/n^2$, $n = 1, 2, \ldots$ führten auf eine Rydberg-Konstante R_H, die mit der aus den gemessenen Wellenlängen gewonnenen in voller Übereinstimmung stand. LYMAN fand 1916 die zu $n = 1$ gehörende Serie im Ultravioletten. BOHR wurde bei seinen Überlegungen, die Plancksche Quantenhypothese auf das Atom zu übertragen, dadurch bestärkt, daß sich aus den Konstanten Masse des Elektrons m, Elementarladung e und Plancksches Wirkungsquantum h eine Kombination der Dimension Länge bilden ließ, deren Größenordnung mit der aus gaskinetischen Betrachtungen bekannten Größe der Atome übereinstimmte. Der *Bohrsche Wasserstoffradius*

$a_0 = \hbar^2/(m\,\epsilon^2) = 0{,}5292 \cdot 10^{-10}\,\text{m}$, $\epsilon^2/a_0 = 27{,}212\,\text{eV}$,

wird daher bei atomaren Betrachtungen als Bezugslänge benutzt, ϵ^2/a_0 als Bezugsgröße für Energien ($\epsilon^2 \equiv e^2/4\pi\varepsilon_0$). Man nennt diese Größen auch *atomare Einheiten*.

Die *Lösung der radialen Schrödinger-Gleichung* führt im Coulomb-Potential auf die Eigenwerte

$$E_n = -\frac{\epsilon^2}{2a_0}\,\frac{1}{(n_r + l + 1)^2} = -\frac{\epsilon^2}{2a_0}\,\frac{1}{n^2}$$

(Abb. 4.2.2-1). Es gehört zu den Besonderheiten des Coulomb-Potentials, daß die Eigenwerte der zu verschiedenen Drehimpulsquantenzahlen l gehörenden Effektivpotentiale miteinander entartet sind (Abb. 4.2.2-2). Zur Nummerierung der Zustände verwendet man daher statt der *Knotenzahl* n_r die *Hauptquantenzahl* n. Anstelle der *Drehimpulsquantenzahlen* $l = 0, 1, 2, 3, \ldots$ werden die Buchstaben s, p, d, f, ... verwendet, so daß man vom 1s-Niveau, 2s-Niveau, 2p-Niveau usw. spricht. Von der Quantenzahl m hängt die Energie nicht ab, im kugelsymmetrischen Potential kann die Energie eben nicht von der *Richtung des Drehimpulses* abhängen. Jedes Niveau ist demzufolge $(2l+1)$-fach entartet, da es $2l+1$ verschiedene m-Werte zu einem l gibt. Bei Berücksichtigung des *Elektronenspins* (vgl. 4.3.) findet man, daß sich die Zahl der miteinander entarteten Zustände verdoppelt, da der Spin zwei Einstellmöglichkeiten hat. Für die *Wellenfunktionen im Coulomb-Potential* (vgl. Abb. 4.2.2-2) erhält man die Laguerreschen Polynome mit

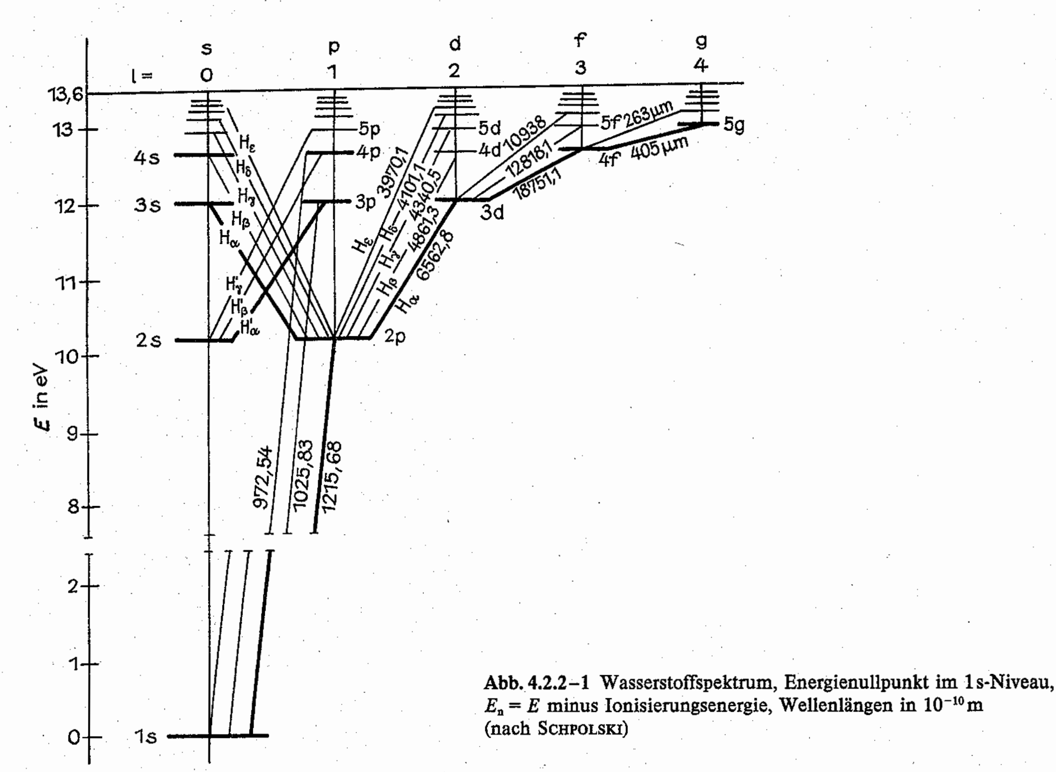


Abb. 4.2.2-1 Wasserstoffspektrum, Energienullpunkt im 1s-Niveau, $E_n = E$ minus Ionisierungsenergie, Wellenlängen in 10^{-10} m (nach SCHPOLSKI)

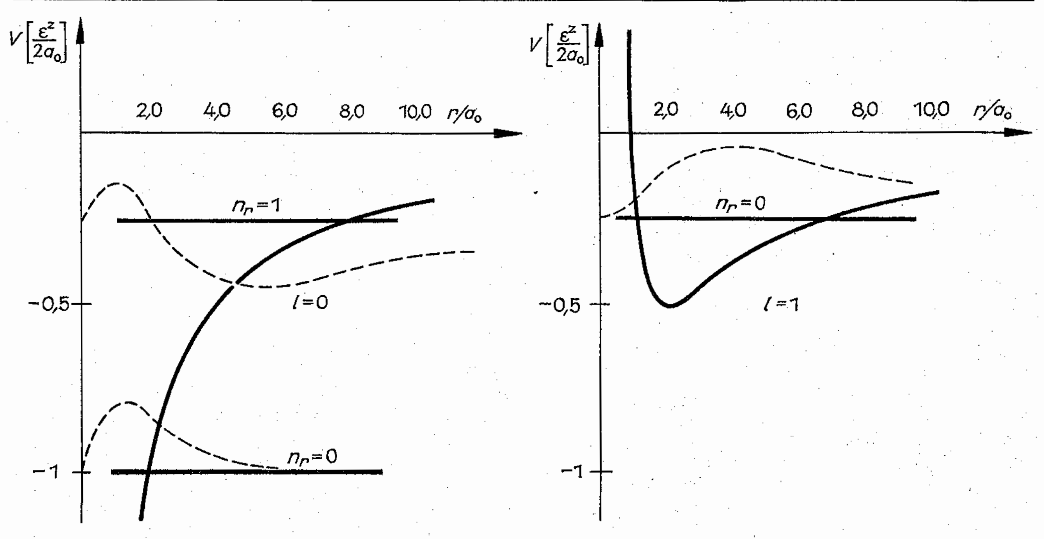


Abb. 4.2.2–2 Entartung der Zustände gleicher Hauptquantenzahl und Radialteil der Wellenfunktion (nicht normiert)

einem Exponentialfaktor multipliziert, $u_{1s} = (2/\sqrt{a_0})$ $(r/a_0)e^{-r/a_0}$, $u_{2s} = (1/\sqrt{2a_0})$ (r/a_0) $(1 - r/2a_0)e^{-r/2a_0}$, $u_{2p} = (1/\sqrt{24a_0})$ $(r/a_0)^2 e^{-r/2a_0}$ usw.
Berücksichtigt man die *Mitbewegung des Kerns*, des Protons im Wasserstoffatom, dann erhält man für die Differenzkoordinate die Bewegung eines Teilchens mit der *reduzierten Masse* $m/(1 + m/M)$ im gleichen Potential. Die Rydberg-Konstante für einen Kern der Masse M und der Kernladungszahl Z berechnet sich daher aus $R_{M,Z} = Z^2 R_\infty/(1 + m/M)$, $R_\infty = 109\,737{,}31\ \mathrm{cm}^{-1}$. Für das Wasserstoffatom ergibt sich $R_H = 109\,677{,}69\ \mathrm{cm}^{-1}$. Die durch die Kernmasse bedingte Wellenlängenänderung führte z. B. zur Entdeckung des Deuteriums. Auch das Verhältnis $m_p/m = R_H/(R_\infty - R_H) = 1\,836{,}15$ läßt sich bestimmen.

4.2.3. Dreidimensionaler harmonischer Oszillator

Das Potential $V(r) = m\omega^2 r^2/2$ führt auf das Eigenwertspektrum $E_{n,l} = \hbar\omega(2n_r + l + 3/2)$, $n_r = 0, 1, 2, \ldots, l = 0, 1, 2, \ldots$, wobei jeder Zustand $2l + 1$-fach entartet ist. Die Energie läßt sich auch als Summe der Energien dreier eindimensionaler Oszillatoren angeben, $E_{n_x n_y n_z} = (n_x + n_y + n_z + 3/2)$, $n_x = 0, 1, 2, \ldots$ usw.

4.2.4. Zweiatomige Moleküle

Bei zweiatomigen Molekülen (Abb. 4.2.4–1) kann man das Effektivpotential zwischen den beiden Atomen in der Umgebung des Minimums r_0 durch eine Parabel annähern, $V_{\mathrm{eff}} = V_l + (m/2)\omega_l^2 (r - r_l)^2$, was auf die Energiewerte $E_{n,l} = -V_l + \hbar\omega_l(n_r + 1/2)$ führt. Drückt man V_l und ω_l durch die Größen für $l = 0$ aus, dann bekommt das Energiespektrum die Gestalt $E_{n,l} = V_0 + \hbar\omega_0(n_r + 1/2) + \hbar^2 l(l + 1)/2\Theta + \Delta E$. Der zweite Summand gibt die *Schwingungsniveaus* an, der dritte entspricht der Energie eines *Rotators* mit dem Trägheitsmoment $\Theta = mr_0^2$. ΔE enthält Korrekturterme, die das Schwingungs- und Rotationsspektrum koppeln und die die Anharmonizität (die Abweichungen von der Parabel) des Potentials berücksichtigen. Da die Rotationsterme wegen des großen Trägheitsmomentes sehr eng im Vergleich zu den Schwingungsniveaus liegen, treten die für Moleküle typischen Bandenspektren auf.

4.3. Drehimpulse

Bahndrehimpuls, Spin, Gesamtdrehimpuls sind physikalische Größen, die zur Klassifizierung der Zustände (z. B. in einem Atom) benutzt werden, weil Drehimpulsausdrücke oft als Erhaltungsgrößen auftreten. Weiterhin ist ein zum Drehimpuls proportionales magnetisches Moment vorhanden. Die Wechselwirkung zwischen diesen magnetischen Momenten bewirkt eine Feinstruktur der Spektralterme, die nur von den Drehimpulsen ausgehend verstanden werden kann. Als wesentlicher Unterschied zur klassischen Mechanik wird eine *Richtungsquantelung* beobachtet: Bezüglich einer vorgegebenen Richtung (z-Richtung) kann die

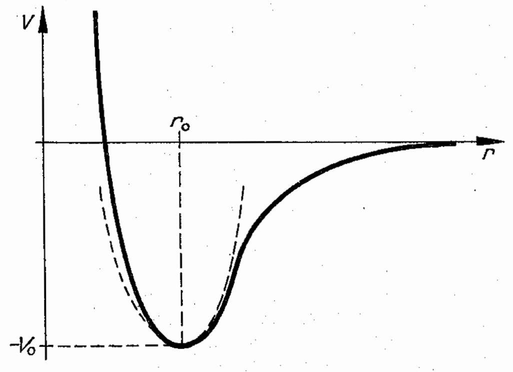


Abb. 4.2.4–1 Ion-Ion-Wechselwirkung im zweiatomigen Molekül und seine Annäherung durch eine Parabel (– – –)

z-Komponente des Drehimpulses J_z nur diskrete Werte annehmen, wobei $\Delta J_z = \hbar$ gilt.

4.3.1. Drehimpulsoperatoren

Wie alle physikalischen Größen werden auch Drehimpulse in der Quantenmechanik durch Operatoren beschrieben. Unter einem Drehimpuls $\boldsymbol{J}$ versteht man einen Operator, dessen Komponenten den *Vertauschungsrelationen*

$$[J_x, J_y] = -(\hbar/\mathrm{i})J_z, \quad [J_y, J_z] = -(\hbar/i)J_x, \quad [J_z, J_x] = -(\hbar/i)J_y$$

genügen. Allein aus diesen Vertauschungsrelationen – und damit für alle Drehimpulse, ob Bahndrehimpuls oder Spin, gültig – kann man folgende Aussagen über Eigenwerte und Eigenfunktionen ableiten:

1. Es gilt $[J^2, J_z] = 0$. Zum Quadrat und zu einer Komponente des Drehimpulses (man greift üblicherweise die z-Komponente heraus) läßt sich ein gemeinsames System von Eigenfunktionen $|jm\rangle$ finden, z. B. die Kugelflächenfunktionen Y_{lm} beim Bahndrehimpuls.
2. J_z hat die Eigenwerte $\hbar m, J_z|jm\rangle = \hbar m|jm\rangle$, wobei $\Delta m = 1$ gilt. Mit j wird der maximale Wert von m bezeichnet $m_{\max} = j$. Da die auftretenden Meßwerte nicht von der Wahl des Richtungssinnes der z-Achse abhängen können, muß neben m auch $-m$ Eigenwert sein. Das ist bei $\Delta m = 1$ nur realisierbar, wenn j ganzzahlig oder halbzahlig ist (Abb. 4.3.1-1). Es gilt also $-j \leqq m \leqq j$ mit $j = 0, 1, 2, \ldots$ oder $j = 1/2, 3/2, 5/2, \ldots$
3. Die Eigenwerte von J^2 sind $\hbar^2 j(j+1)$, $J^2|jm\rangle = \hbar^2 j(j+1)|jm\rangle$.
4. Die über $J_\pm = J_x \pm \mathrm{i}J_y$ definierten Operatoren sind *Umklappoperatoren.* Aus einem Zustand $|jm\rangle$ mit dem Eigenwert $\hbar m$ von J_z erzeugt J_+ einen Zustand mit dem Eigenwert $\hbar(m+1)$, J_- einen mit dem Eigenwert $\hbar(m-1)$, $J_\pm|jm\rangle = \hbar\sqrt{j(j+1) - m(m \pm 1)}\,|j, m \pm 1\rangle$. Da j der maximale Wert von m ist, gilt $J_+|jj\rangle = 0$, ebenso $J_-|j-j\rangle = 0$.

4.3.2. Elektronenspin

Mit der Beobachtung der Dublettstruktur bestimmter Spektralterme und des anomalen Zeeman-Effektes traten in den zwanziger Jahren Probleme bei der Klassifizierung der Atomniveaus auf, die Uhlenbeck und Goudsmit 1925 mit der Hypothese klären konnten, daß das Elektron einen *Eigendrehimpuls* – den Elektronenspin – mit zwei Einstellmöglichkeiten $m_s = \pm 1/2$ hat. Mit dem *Stern-Gerlach-Versuch* wurde 1927 diese Hypothese experimentell bestätigt. Eine rotierende Ladung $(-e)$ oder ein Ringstrom hat ein magnetisches Moment $\mu = g(-e/2m)J$, das zum Drehimpuls proportional ist. Da J von der Größenordnung $\hbar$ ist, führte man als Bezugsgröße das *Bohrsche Magneton* $\mu_B = e\hbar/2m = 9{,}274 \cdot 10^{-24}\,\mathrm{Am}^2$ ein. Elektrodynamik und klassische Mechanik liefern für den *gyromagnetischen Faktor* $g = 1$. Dieser Wert wurde in der Quantenmechanik auch für das mit dem Bahndrehimpuls verknüpften magnetischen Moment bestätigt, $g_L = 1$.

Stern und Gerlach schickten einen Atomstrahl (Ag-Atome, das Experiment läuft in gleicher Weise für H-Atome) durch ein inhomogenes Magnetfeld $\boldsymbol{B}$, in dem Kräfte auf einen magnetischen Dipol wirken. Sie beobachteten eine Aufspaltung des Strahles in zwei Teilstrahlen. Da das Elektron im Wasserstoffatom im Grundzustand den Bahndrehimpuls Null hat, konnte das beobachtete magnetische Moment nur zu einer Eigenrotation des Elektrons gehören. Da die Kräfte auf den magnetischen Dipol von dessen Orientierung im Magnetfeld abhängen, konnte dieses magnetische Moment μ_S bzw. der Eigendrehimpuls $\boldsymbol{\check{s}}$ nur zwei Einstellmöglichkeiten haben. Der Beweis für die Existenz des Elektronenspins und für dessen Halbzahligkeit war erbracht worden. Aus der Größe der Aufspaltung des Atomstrahles folgte für $\mu_{S,z} = \pm\mu_B$, also ein zum Elektronenspin gehörender gyromagnetischer Faktor $g_S = 2$. Er ist klassisch nicht erklärbar. Die relativistische Quantenmechanik liefert den richtigen Faktor. Der *Spinoperator* $\boldsymbol{\check{s}}$ ist ein Drehimpulsoperator. Von den Drehimpulsquantenzahlen j, m (vgl. 4.3.1.) hat für ein einzelnes Elektron j immer den Wert 1/2, während m (für den Spin m_S genannt) die beiden Werte $m_S = \pm 1/2$ annehmen kann. j wird daher nicht explizit in den Zuständen aufgeführt.

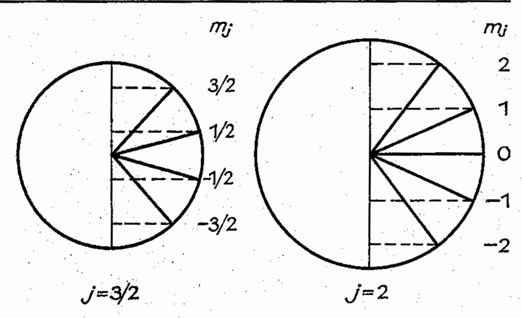


Abb. 4.3.1-1 Quantelung des Drehimpulses. Der Kreisradius entspricht $\sqrt{j(j+1)}$

Das Elektron hat vier Freiheitsgrade. Neben dem Ortsvektor tritt in seiner Wellenfunktion eine Spinvariable s auf. Dies ist eine diskrete Variable, die nur die beiden Werte $s = \pm 1/2$ annehmen kann. Nennt man den Spinanteil der Wellenfunktion $\chi(s)$, dann ist $|\chi(1/2)|^2$ die Wahrscheinlichkeit dafür, das Elektron mit der z-Komponente des Spins $+\hbar/2$ anzutreffen. Die *Eigenfunktionen* von $\check{s}_z$ sind $\chi_{m_S}(s) = \delta_{m_S,s}$, $\check{s}_z\chi_{m_S}(s) = \hbar m_s \chi_{m_S}(s)$. Da nur zwei Eigenfunktionen auftreten, kurz χ_+ bzw. χ_- genannt, ist die Matrixschreibweise für Operatoren und Eigenfunktionen zweckmäßig, $\chi_+ = \begin{pmatrix}1\\0\end{pmatrix}$ bzw. $\chi_- = \begin{pmatrix}0\\1\end{pmatrix}$. Führt man über $\boldsymbol{\check{s}} = (\hbar/2)\boldsymbol{\sigma}$ die dimensionslosen, sogenannten *Pauli-Matrizen* ein, dann wird der Spinoperator durch die Zwei-Mal-Zwei-Matrizen

$$\sigma_x = \begin{pmatrix}0 & 1\\1 & 0\end{pmatrix}, \quad \sigma_y = \begin{pmatrix}0 & -i\\1 & 0\end{pmatrix}, \quad \sigma_z = \begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$$

beschrieben. Die Pauli-Matrizen haben die Eigenschaften $\sigma = \sigma_i^+$, $\sigma_i^2 = E$ (Einheitsmatrix), $\sigma_x\sigma_y = i\sigma_z$ und $\sigma_x\sigma_y = -\sigma_y\sigma_x$.

4.3.3. Addition von Drehimpulsen

Zwei Drehimpulse $\boldsymbol{J}_1$ bzw. $\boldsymbol{J}_2$ mit den Eigenzuständen $|j_1m_1\rangle$ bzw. $|j_2m_2\rangle$ mögen sich zu einem *Gesamtdrehimpuls* $\boldsymbol{J} = \boldsymbol{J}_1 + \boldsymbol{J}_2$ zusammensetzen. Da in der Quantenmechanik nur Quadrat- und z-Komponente gleichzeitig meßbar sind, kann man die Meßwerte j und m nicht – wie in der klassischen Mechanik – durch eine Vektorsumme ermitteln. Die Quantenzahl j beschreibt die maximale z-Komponente von $\boldsymbol{J}$. Ihr größter Wert ist $j_1 + j_2$. Insgesamt kann sie die Werte

$|j_1 - j_2| \leq j \leq j_1 + j_2$ durchlaufen, was klassisch dem Übergang von „Antiparallelstellung" zur „Parallelstellung" entspricht. Bei gegebenem j gibt es für m die $2j+1$-Werte $-j \leq m \leq j$. Das Produkt der Eigenfunktionen $|j_1 m_1\rangle |j_2 m_2\rangle$ ist eine Eigenfunktion von $\boldsymbol{J}_1^2, J_{1z}, \boldsymbol{J}_2^2, J_{2z}$. Ebenso ist es eine Eigenfunktion von J_z mit dem Eigenwert $\hbar(m_1 + m_2)$, nicht aber eine Eigenfunktion von $\boldsymbol{J}^2$. Da in $\boldsymbol{J}^2 = \boldsymbol{J}_1^2 + \boldsymbol{J}_2^2 + 2\boldsymbol{J}_1 \cdot \boldsymbol{J}_2$ mit $\boldsymbol{J}_1 \cdot \boldsymbol{J}_2 = J_{1x}J_{2x} + J_{1y}J_{2y} + J_{1z}J_{2z}$ alle Komponenten von $\boldsymbol{J}_1$ und $\boldsymbol{J}_2$ explizit auftreten, können $\boldsymbol{J}^2$ und J_{1z} bzw. J_{2z} nicht gleichzeitig gemessen werden, wohl aber $\boldsymbol{J}^2, \boldsymbol{J}_1^2, \boldsymbol{J}_2^2, J_z$. Die Eigenfunktionen $|jj_1 j_2 m_j\rangle$ dieser Operatoren sind Linearkombinationen zu verschiedenen Werten von m_1 und m_2 mit der Einschränkung $m_1 + m_2 = m$. Betrachtet man Bahn- und Spindrehimpuls ($j_2 = 1/2$) eines Elektrons, dann erhält man

$$\left| jl\frac{1}{2}m_j \right\rangle = \frac{1}{\sqrt{2l+1}} \left[\sqrt{l + \frac{1}{2} \pm m_j}\; Y_{l,\, m_j - 1/2}\chi_+ \mp \sqrt{l + \frac{1}{2} \mp m_j}\; Y_{l,\, m_j + 1/2}\chi_- \right]$$

für $j = l \pm 1/2$ als Eigenfunktionen von $\boldsymbol{J}^2, \boldsymbol{L}^2, \boldsymbol{s}^2$ und J_z.

4.3.4. Spinbahnkopplung

Die mit der Bahnbewegung bzw. der Eigenrotation verknüpften magnetischen Momente $\boldsymbol{\mu}_L \sim \boldsymbol{L}$ bzw. $\boldsymbol{\mu}_S \sim \boldsymbol{s}$ haben eine Wechselwirkungsenergie, die zu einem Zusatzterm $H_{LS} = \frac{1}{4}\left(\frac{\hbar}{mc}\right)^2 \frac{1}{r}\frac{dV}{dr}\frac{2\boldsymbol{L}\cdot\boldsymbol{s}}{\hbar^2}$, der sogenannten Spinbahnkopplung, im Hamilton-Operator $H = p^2/2m + V(r) + H_{LS}$ führt. Die Energie der Eigenzustände, z.B. des Wasserstoffatoms, sind von der Orientierung der magnetischen Momente zueinander bzw. der Drehimpulse zueinander abhängig. Dies führt zu der sogenannten *Feinstruktur* des Termschemas (vgl. 4.7.2.).

4.4. Streuprobleme

Beim Streuproblem fällt ein Teilchenstrom *bekannter Energie* und der Stromdichte $j_0 = \Delta N_0/\Delta A\,\Delta t$ auf einen Streuer (vgl. 2.1.10.). Man hat also kein Eigenwertproblem zu lösen. Die Anzahl der in den Raumwinkel $d\Omega = \sin\vartheta\, d\vartheta\, d\varphi$ gestreuten Teilchen $dN = \sigma j_0 d\Omega \Delta t$ wird gemessen. Der auftretende Proportionalitätsfaktor $\sigma(\Omega, E)$ wird als *differentieller Wirkungsquerschnitt* bezeichnet, der *Gesamtwirkungsquerschnitt* $\sigma(E) = \int d\Omega\, \sigma(\Omega, E)$ ergibt sich durch Integration über alle Winkel. Im Unterschied zur klassischen Mechanik ist der Wirkungsquerschnitt nicht aus der Bahnkurve, sondern aus der Lösung der Schrödinger-Gleichung zu gewinnen.

4.4.1. Tunneleffekt

Beim eindimensionalen Streuproblem treten nur Durchlaß- und Reflexionskoeffizient auf (Abb. 4.4.1–1). Für ein klassisches Teilchen gilt $|R|^2 = 1, D = 0$ für $E < V_0$ bzw. $R = 0, |D|^2 = 1$ für $E > V_0$. Wie Abb. 4.4.1–2 zeigt, ist die quantenmechanische Lösung davon völlig verschieden. Insbesondere kann ein Teilchen mit $E < V_0$ die Potentialschwelle mit einer – wenn auch kleinen – Wahrscheinlichkeit durchtunneln. Der *Durchlaßkoeffizient* verhält sich wie $|D|^2 \sim e^{-2G}$, worin der *Gamov-Faktor*

$$G = \int_{x_-}^{x_+} dx \sqrt{\frac{2m}{\hbar^2}[V(x) - E]}$$

durch Höhe und Breite der Schwelle bestimmt wird.
Beim *α-Zerfall* durchtunnelt ein Heliumkern $Z = 2$, Masse m, der durch die Kernkräfte im Innern des Atomkernes gehalten wird, die durch die Coulomb-Abstoßung bedingte Potentialschwelle (Abb. 4.4.1–3). Für kleine Energien E des α-Teilchens ist der Gamov-Faktor ungefähr $G = Zz\, e^2/(\hbar c)\sqrt{2mc^2/E}\,\pi/2$. Für die Kernladungszahl $Z = 90$ ergibt sich $G \approx 150/\sqrt{E/\text{MeV}}$. Im Bereich von $E = 4 \ldots 9$ MeV variiert G zwischen 75 und 50. Der Durchlaufkoeffizient ändert sich in diesem Bereich um einen Faktor e^{50}. Daraus werden die großen Unterschiede und die starke Energieabhängigkeit der Halbwertszeit bzw. Lebensdauer verschiedener α-Strahler verständlich.
Ein Anfangszustand, wie der eines Teilchens im Innern des Potentials von Abb. 4.4.1–3 mit $E > 0$, ist kein stationärer Zustand, da das Teilchen die Schwelle durchtunneln kann. Man nennt ihn auch einen *virtuell gebundenen Zustand*. Um einen solchen lokalisierten Anfangszustand aufzubauen, muß man stationäre Lösungen aus einem Energiebereich ΔE überlagern. Dieser lokalisierte Zustand hat also eine gewisse Energiebreite, die nur für unendlich lange Lebensdauer τ, d. h. für undurchdringbare Schwelle, gegen Null geht (Abb. 4.4.1–4). Es gilt die *Energie-Zeit-Unschärferelation* $\Delta E \cdot \tau \geq \hbar/2$. Ein Zustand mit endlicher Lebensdauer hat eine aus dieser Relation folgende Linienbreite.

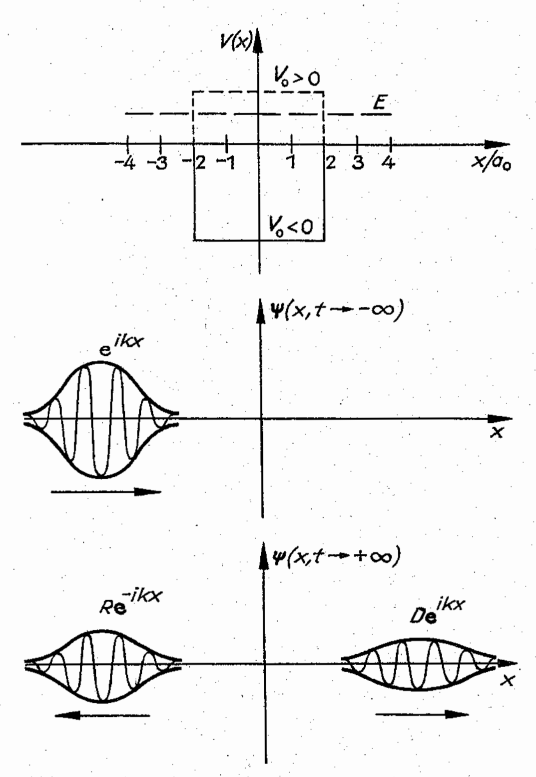


Abb. 4.4.1–1 Eindimensionales Streuproblem

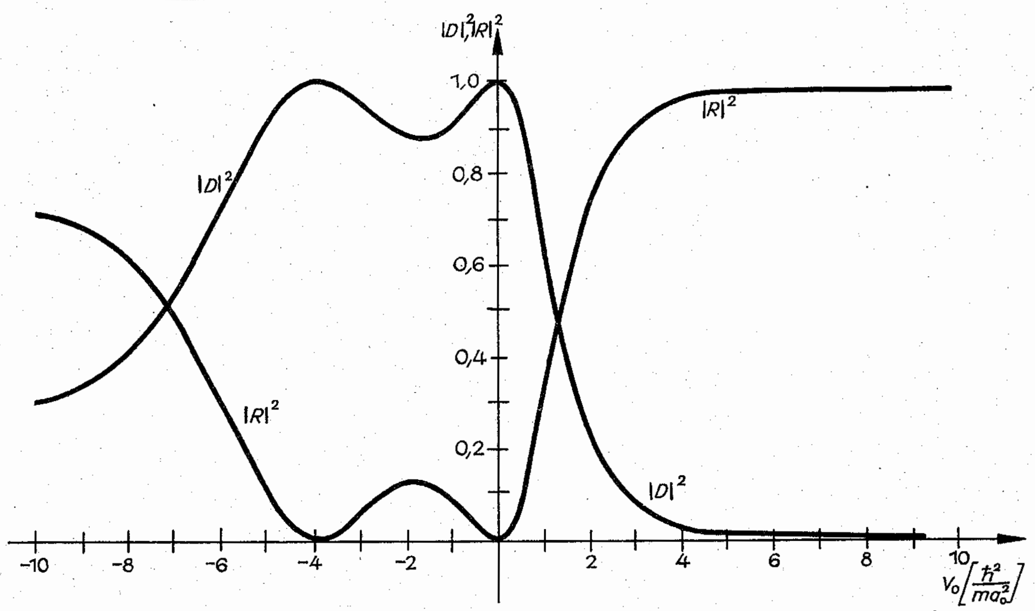


Abb. 4.4.1–2 Abhängigkeit von $|R|^2$ und $|D|^2$ von der Tiefe V_0 des Potentiales aus Abb. 4.4.1–1 für $E = \hbar^2/(ma_0^2)$

Abb. 4.4.1–3 Potentialverlauf beim α-Zerfall von ${}^{92}U^{238} \to {}^{90}Th^{234}$ (nach HERTZ)

Abb. 4.4.1–4 Anfangszustand (– – –) als (a) gebundener bzw. (b) und (c) als virtuell gebundener Zustand endlicher Energiebreite

4.4.2. Streuamplitude

Das quantenmechanische Streuproblem besteht im Aufsuchen derjenigen Lösung der Schrödinger-Gleichung für ein Teilchen der Energie $E = \hbar^2k^2/2m$, die – wie aus der Problemstellung folgt – asymptomatisch, d. h. in großen Abständen vom Streuzentrum, aus der einfallenden ebenen Welle $\Psi_0 = e^{ik\cdot r}$ und aus einer auslaufenden Kugelwelle, der Streuwelle, besteht,

$$[p^2/2m + V(r)]\Psi(r) = \hbar^2k^2/2m\Psi(r),$$

$$\Psi_{\text{asympt.}} = \Psi_0 + f(\Omega, k)\frac{e^{ikr}}{r}.$$

Die Streuamplitude $f(\Omega, k)$ beschreibt den Teilchenstrom in Richtung Ω. Aus ihr kann der differentielle Wirkungsquerschnitt berechnet werden $\sigma(\Omega, E) = |f(\Omega, k)|^2$. Das *optische Theorem* $\sigma(E) = (4\pi/k) Im f(0, k)$, zeigt, daß die Gesamtzahl der gestreuten Teilchen berechnet werden kann, indem man betrachtet, was im durchgehenden Teilchenstrom fehlt. Die Streuamplitude läßt sich aus $f(\Omega, k) = -(2m/\hbar^2)(1/4\pi)\int d\tau e^{-ik'\cdot r}V(r)\Psi(r)$ berechnen. Sie setzt sich aus Beiträgen zusammen, die aus den einzelnen Volumenelementen des Streuers stammen und mit verschiedenen Phasenfaktoren $e^{-k'r}$ überlagert werden. k' zeigt in Richtung Ω.

4.4.2.1. Streuphasen charakterisieren den asymptomatischen Verlauf der Wellenfunktion beim kugelsymmetrischen Streuer. Die radiale Schrödinger-Gleichung hat im

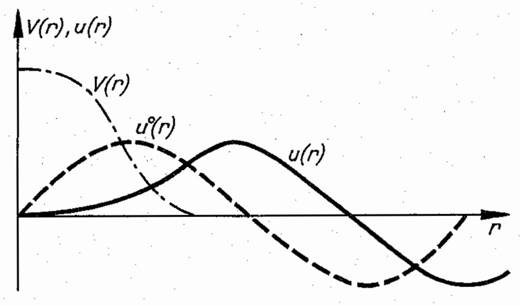


Abb. 4.4.2–1 Phasenverschiebung durch Streupotential

konstanten Potential $V = 0$ sphärische Bessel-Funktionen $u_0(r) = krj_l(kr)$ als Lösung. Sie hängen mit den in der Mathematik üblichen Bessel-Funktionen über $j_l(x) = \sqrt{\pi/2x}\, I_{l+1/2}(x)$ zusammen. $krj_l(kr) \to \sin(kr - l\pi/2)$ ist ihr asymptotisches Verhalten. Im Streupotential $V(r)$ ist der Radialteil außerhalb des Streuers nur phasenverschoben (Abb. 4.4.2–1), verläuft also asymptotisch wie $u(r) \to \sin(kr - l\pi/2 + \eta_l)$. Die Streuphasen $\eta_l(E)$ charakterisieren jetzt den asymptotischen Verlauf der Wellenfunktion. Demzufolge lassen sich auch die Streuamplitude und der Wirkungsquerschnitt aus ihnen berechnen,

$$f(\vartheta, k) = 1/k\sum_l (2l+1)e^{i\eta_l}\sin\eta_l P_l(\cos\vartheta),$$

$$\sigma(\vartheta, E) = 4\pi/k^2\sum_l (2l+1)\sin^2\eta_l.$$

Bei begrenzter Reichweite des Potentials $V(r)$ wird dies im Effektivpotential mit wachsendem l stärker und stärker vom Zentrifugalpotential überdeckt. Daher werden die Streuphasen η_l mit wachsendem l rasch klein, und es genügen wenige l-Summanden. Das ist der Vorteil der Drehimpulsentwicklung. Bei der niederenergetischen Streuung von Nukleonen (bis ≈ 10 MeV) ist der Term $l = 0$ ausreichend. Die Streuung der Elektronen an den Ionenpotentialen in einem Metall kann für Energien bis zur Fermi-Energie mit den Streuphasen η_0, η_1, η_2 ausreichend erfaßt werden.

4.4.2.2. Resonanzstreuung. Einen großen Beitrag zum Wirkungsquerschnitt liefert eine Streuphase, die in der Nähe von $\pi/2$ liegt. Dies ist in der Umgebung der Resonanzenergie E_r der Fall (vgl. Abb. 4.4.1–4). Bei der Streuung mit $E \approx E_r$ erhält man eine Wellenfunktion, deren Amplitude im Innern der Schwelle groß gegen die der einfallenden Welle ist. Dies tritt jedoch nur in einem engen Energiebereich auf, dessen Breite durch den Durchlaßkoeffizienten der Schwelle bestimmt ist. In diesem Energiebereich vergrößert sich die Streuphase um π, durchläuft also den Wert $\pi/2$.

4.5. Störungsrechnung und ihre Anwendung

Die Schrödinger-Gleichung läßt sich nur in wenigen Fällen analytisch lösen, die konkrete Auswertung muß in der Regel numerisch erfolgen. Geeignete *Näherungsmethoden* sind dabei gefragt. Wird ein System schwach gestört, dann stellt sich die Störungsrechnung die Aufgabe, die Abweichungen der Eigenwerte $E_n - E_n^0$ und Eigenfunktionen $\varphi_n - \varphi_n^0$ von denen des ungestörten Systems $H^0\varphi_n^0 = E_n^0\varphi_n^0$ nach Potenzen des *Störpotentials* V zu entwickeln. Hält man sich vor Augen, daß das elektrische Feld eines Protons im Abstand Bohrscher Wasserstoffradius die – für makroskopische Begriffe – ungeheuer große Stärke $e/4\pi\varepsilon_0 a_0^2 \approx 5\cdot 10^{11}$ V/m aufweist, dann versteht man, daß angelegte Felder für atomare Systeme als schwache Störung verstanden werden können.

Die Entwicklung von Näherungsmethoden wird dadurch erleichtert, daß der Erwartungswert $\bar{E}[\varphi] = \int d\tau\varphi^* H\varphi$ – als Funktional von φ betrachtet – ein Minimum für $\varphi = \varphi_0$ hat. Für $\varphi = \varphi_n$ verfügt er über einen Sattelpunkt, der einem Minimum entspricht, wenn man nur Funktionen φ betrachtet, die orthogonal zu $\varphi_0, \ldots, \varphi_{n-1}$ sind. Berechnet man $\bar{E}$ mit einer Näherungsfunktion für den Grundzustand, dann erhält man einen Näherungswert für die Grundzustandsenergie E_0 mit einem Fehler zweiter Ordnung bezüglich $\varphi - \varphi_0$. Außerdem liegt dieser Näherungswert immer über E_0.

Darauf baut das *Ritzsche Variationsprinzip* auf. Zur Berechnung von E_0 wird für φ_0 eine Funktion $f(\alpha, \beta)$ angesetzt, die den bekannten qualitativen Verlauf von φ_0 hat, aber noch offene Parameter $\alpha, \beta, \ldots$ (die z. B. Breite und Höhe bestimmen) enthält. Diejenige Funktion f beschreibt φ_0 am besten, für die der Erwartungswert $\bar{E}(\alpha, \beta)$ ein Minimum aufweist $\partial\bar{E}/\partial\alpha = 0, \partial\bar{E}/\partial\beta = 0$. Die daraus resultierenden Parameterwerte α_0, β_0 liefern den besten Näherungswert $\bar{E}(\alpha_0, \beta_0) \gtrsim E_0$ für die Grundzustandsenergie.

Betrachtet man ein System (H_0), das einer *schwachen Störung* V unterworfen wird, d. h. den Hamilton-Operator $H = H^0 + V$ hat, dann sind die ungestörten Lösungen φ_n^0 Näherungsfunktionen für die gestörten Lösungen φ_n mit einem Fehler in erster Ordnung in V. Die mit φ_n^0 berechneten Eigenwerte $E_n = E_n^0 + V_{nn} + 0(V^2)$ enthalten den Einfluß von V in erster Ordnung richtig. Als Zusatzterm tritt das Matrixelement der Störung V_{nn} auf,

$$\left(V_{mn} \equiv \int d\tau\varphi_m^* V\varphi_n\right).$$

4.5.1. Zeitunabhängige Störungsrechnung

In ihr werden die Korrekturen zu E_n^0 systematisch als Beiträge in verschiedenen Ordnungen der Störung $V(r)$ berechnet. Das Ergebnis ist

$$E_n = E_n^0 + V_{nn} - \sum_{m(\neq n)} \frac{|V_{mn}|^2}{E_m^0 - E_n^0} + \ldots,$$

$$\varphi_n = \varphi_n^0 - \sum_{m(\neq n)} \varphi_m^0 \frac{V_{mn}}{E_m^0 - E_n^0} + \ldots .$$

Man beachte, daß φ_n nicht auf Eins normiert ist, $\int \mathrm{d}\tau \varphi_n^{0*} \varphi_n = 1$. Die Summen werden nur geringe Beiträge liefern, wenn die Matrixelemente V_{mn} klein gegen die Niveauabstände $E_m^0 - E_n^0$ sind. Im anderen Falle ist die Störungsrechnung in der Regel nicht anwendbar.

Gesondert muß der Fall betrachtet werden, in dem die *ungestörten Niveaus entartet* sind, z.B. $E_1^0 = E_2^0 \equiv E^0$ gilt. In diesem Falle ist jede Linearkombination $\varphi^0 = c_1\varphi_1^0 + c_2\varphi_2^0$ eine ungestörte Lösung mit der Energie E^0. In die Störungsreihe sind diejenigen Linearkombinationen einzusetzen, in die die Zustände φ_1 bzw. φ_2 für verschwindende Störung $V \to 0$ übergehen. In den Summen treten dann die Terme mit gleichen Energien in den Nennern nicht mehr auf. Geht man von zwei beliebigen Funktionen aus, dann erhält man für die Energie in erster Störungsordnung $E_{1,2} = E^0 + (V_{11} + V_{22})/2 \pm \sqrt{(V_{11} - V_{22})^2 + 4|V_{12}|^2}$. Bei dem z. B. bei der Berechnung von Energiebändern anzutreffenden Fall $V_{11} = V_{22} = V_0$ vereinfacht sich diese Formel zu

$$E_{1,2} = E^0 + V_0 \pm |V_{12}|.$$

4.5.2. Zeitabhängige Störungsrechnung

Ist die Störung $V(t)$ zeitabhängig, dann wird der Hamilton-Operator zeitabhängig. Die Energie ist demzufolge keine Erhaltungsgröße mehr. Typische Prozesse in zeitabhängigen Feldern sind daher *Übergänge zwischen verschiedenen Energieniveaus*, Absorptions- und Emissionsvorgänge.

Entwickelt man die Wellenfunktion $\Psi(t) = \sum_n c_n(t)\varphi_n^0 \mathrm{e}^{-(\mathrm{i}/\hbar)E_n^0 t}$ nach den Lösungen des ungestörten Systems, dann gilt zum Anfangszeitpunkt $c_i(0) = 1$, $c_n(0) = 0$ für $n \neq i$, wenn sich das System vor Einwirken der Störung im Zustand φ_i^0 befunden hat. $|c_f(t)|^2$ gibt die Wahrscheinlichkeit an, das System zur Zeit t im Zustand φ_f anzutreffen. Für eine rein periodische Störung $V(t) = A\,\mathrm{e}^{-\mathrm{i}\omega t} + A^+\mathrm{e}^{\mathrm{i}\omega t}$ ist die Auswertung am einfachsten. In einem mit ω periodischen Feld kann ein System Quanten der Energie $\hbar\omega$ absorbieren oder emittieren, $E_f = E_i \pm \hbar\omega$. Infolge der endlichen Lebensdauer angeregter Niveaus haben diese eine *natürliche Linienbreite*. Daher werden die Quanten aus einem gewissen Energiebereich emittiert bzw. absorbiert. Die relative Linienbreite liegt für Atomniveaus in der Größenordnung 10^{-8}. Für die Wahrscheinlichkeit je Zeit, daß ein Quant $\hbar\omega$ absorbiert wird, erhält man in erster Ordnung der Störungsrechnung die „goldene Regel“

$$P_{i\to f} = (2\pi/\hbar)\,|A_{fi}|^2\,\delta(E_f - E_i - \hbar\omega)$$

und einen analogen Ausdruck für die Emission. Oft liegt der Endzustand in einem durch eine Zustandsdichte $D(E)$ charakterisierten Kontinuum, dann gilt

$$P_{i\to f} = (2\pi/\hbar)\,|A_{fi}|^2\,D(E_i + \hbar\omega).$$

4.5.3. Emission von Photonen

Eine ebene elektromagnetische Welle wird durch das Vektorpotential

$$A(r,t) = \sqrt{\hbar/(2\varepsilon_0\Omega\omega)}\;\boldsymbol{u}\,(\mathrm{e}^{\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)} + \mathrm{e}^{-\mathrm{i}(\boldsymbol{k}\cdot\boldsymbol{r}-\omega t)})$$

und das skalare Potential $U = 0$ beschrieben. $\boldsymbol{u}$ ist der Polarisationsvektor des elektrischen Feldes. Der Vorfaktor wurde so gewählt, daß die elektromagnetische Energie im Normierungsvolumen Ω den Wert $\hbar\omega$ hat. Aus dem Hamilton-Operator $H = (\boldsymbol{\hat{p}} - q\boldsymbol{A})^2/2m + V(\boldsymbol{r}) + qU = H^0 + H'$ (q ist die Ladung des betrachteten Teilchens) folgt, daß die Störung H' durch das elektromagnetische Feld in erster Ordnung durch $H'(t) = -(q/m)\boldsymbol{A}\cdot\boldsymbol{\hat{p}}$ gegeben ist (mit div $\boldsymbol{A} = 0$). Ist der Anfangs- oder Endzustand des Systems ein am Atom lokalisierter Zustand, dann kann im optischen Bereich wie auch für weiche Röntgenstrahlen der Faktor $\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}$ des Vektorpotentials durch Eins angenähert werden, weil die Wellenlänge groß gegen die Atomausdehnung und damit im Integrationsbereich $\boldsymbol{k}\cdot\boldsymbol{r} \ll 2\pi$ ist. Man nennt dies die *Dipolnäherung*. Mittels $\boldsymbol{\hat{p}}/m = (\mathrm{i}/\hbar)\,[H, \boldsymbol{r}]$ kann das Matrixelement mit dem Impulsoperator in ein solches mit $\boldsymbol{r}$ überführt werden. Für die Wahrscheinlichkeit pro Zeit, daß ein Elektron in einem Atom von einem Zustand φ_i^0 in einen Zustand φ_f^0 übergeht und dabei ein Photon mit der Polarisation $\boldsymbol{u}$ und der Energie $\hbar\omega = E_i - E_f$ in den Raumwinkel $\mathrm{d}\Omega$ emittiert, erhält man aus der goldenen Regel $P\mathrm{d}\Omega = [\epsilon^2/(2\pi\hbar c^3)]\omega^3 \left|\int \mathrm{d}\tau \varphi_f^{0*}\,\boldsymbol{u}\cdot\boldsymbol{r}\varphi_i^0\right|^2 \mathrm{d}\Omega$.

Das Matrixelement enthält *Auswahlregeln*, die die Drehimpulse von Anfangs- und Endzustand verknüpfen: 1. $\Delta l = \pm 1$. Die Lyman-Serie des Wasserstoffatoms, die Übergänge in das 1s-Niveau enthält, ergibt sich also mit 2p, 3p usw. als Anfangszuständen. Das Niveau 2s scheidet als Anfangszustand aus, weil die Übergangswahrscheinlichkeit verschwindet. 2. $\Delta m = 0; \pm 1$. Zu verschiedenen Polarisationsrichtungen $\boldsymbol{u}$ tragen diese drei Möglichkeiten mit verschiedenem Gewicht bei. 3. $\Delta m_s = 0$. Die Orientierung des Spins ändert sich nicht. Dies folgt aus dem in P schon ausgewerteten Spinanteil der Wellenfunktion.

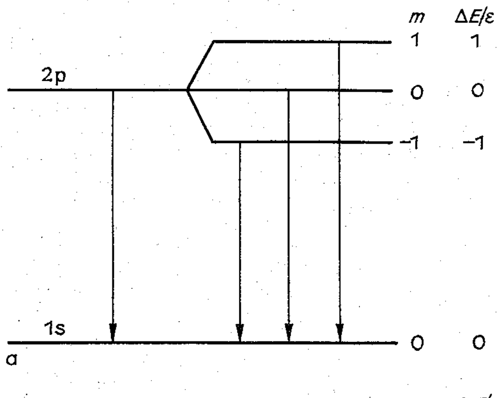
4.5.4. Zeeman-Effekt

Als Zeeman-Effekt bezeichnet man die *Aufspaltung der Spektrallinien im Magnetfeld*. Berücksichtigt man die Spinbahnkopplung $H_{LS} = \frac{1}{4}\left(\frac{\hbar}{mc}\right)^2 \frac{1}{r}\frac{\mathrm{d}V}{\mathrm{d}r}\frac{2\boldsymbol{L}\cdot\boldsymbol{S}}{\hbar^2}$, dann zerlegt man den Hamilton-Operator zweckmäßigerweise in drei Terme $H = H^0 + H_{LS} + H_m$. H^0 enthält ein Zentralpotential und führt auf Energieniveaus E_{nl}, die von der Hauptquantenzahl n und der Nebenquantenzahl l abhängen, beim Wasserstoffatom nur von n. $H^0 + H_{LS}$ liefert die Feinstruktur E_{njl}, beim Wasserstoffatom (vgl. 4.6.2.) z.B. die Aufspaltung des 2p-Niveaus in $2p_{1/2}$ und $2p_{3/2}$ entsprechend den unterschiedlichen Werten des Gesamtdrehimpulses j. H_m beschreibt die Wechselwirkung der mit der Bahn- und Spinbewegung verknüpften magnetischen Momente mit dem äußeren Magnetfeld. Legt man die z-Richtung in Richtung von $\boldsymbol{B} = B\boldsymbol{e}_z$, dann wird $H_m = -\boldsymbol{\mu}\boldsymbol{B} = \mu_B B(L_z + 2\hat{s}_z)/\hbar$. Wegen der unterschiedlichen gyromagnetischen Faktoren tritt vor der z-Komponente des Spinoperators der Faktor 2 auf.

Zur einfacheren Auswertung betrachtet man Spezial- bzw. Grenzfälle. Bei Systemen mit Spin Null tritt keine Spinbahnkopplung auf und kein Beitrag von dem magnetischen

Moment der Spinbewegung. Man spricht in diesem Fall vom normalen Zeeman-Effekt. Bei Systemen mit von Null verschiedenem Spin spricht man in dem Grenzfall, daß die durch die Spinbahnkopplung bewirkte Aufspaltung klein gegen die Aufspaltung durch das äußere Magnetfeld ist, vom Paschen-Back-Effekt, im umgekehrten Fall vom anomalen Zeeman-Effekt.

4.5.4.1. Normaler Zeeman-Effekt. Bei ihm sind die Eigenfunktionen von H^0 mit dem Drehimpulsanteil $|lm\rangle$ auch Eigenfunktionen von $H = H^0 + \mu_B BL_z/\hbar$. $L_z|lm\rangle = \hbar m|lm\rangle$ liefert die Eigenwerte $E_{nlm} = E_{nl} + \mu_B Bm$ (Abb. 4.5.4.1–a). Die möglichen Übergänge ergeben sich unter Beachtung der Auswahlregeln $\Delta l = \pm 1, \Delta m = 0, \pm 1$. Jede Spektrallinie spaltet dreifach auf $\hbar\omega = \hbar\omega_0$; $\hbar\omega_0 \pm \mu_B B$. Man versteht aus diesen Betrachtungen, warum man m auch als magnetische Quantenzahl bezeichnet. Beim normalen Zeeman-Effekt kann man zu bestimmten m-Werten gehörige Übergänge beobachten, damit also auch die Abhängigkeit der Übergangswahrscheinlichkeit von der Polarisationsrichtung feststellen. Parallel zu $\boldsymbol{B}$ werden nur die zwei Linien $\hbar\omega_0 \pm \mu_B B$ emittiert. Sie sind zirkular polarisiert. Senkrecht zu $\boldsymbol{B}$ treten alle drei Linien auf. Sie sind linear polarisiert, für $\hbar\omega_0$ liegt $\boldsymbol{u}$ parallel zu $\boldsymbol{B}$, für $\hbar\omega_0 \pm \mu_B B$ senkrecht zu $\boldsymbol{B}$.



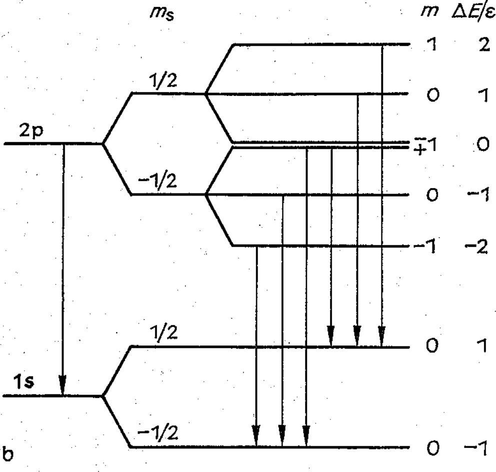


Abb. 4.5.4–1 Dreifache Aufspaltung einer Spektrallinie (a) beim normalen Zeeman-Effekt, (b) Paschen-Back-Effekt ($\varepsilon = \mu_B B$)

4.5.4.2. Paschen-Back-Effekt. Hier wird der Grenzfall $\bar{H}_{LS} \approx 0$ betrachtet. Der Drehimpulsanteil der Eigenfunktionen ist wie für $H^0|lmm_s\rangle$. Mit ihm ergeben sich die Eigenwerte $E_{nlmm_S} = E_{nl} + \mu_B B(m + 2m_S)$ (Abb. 4.5.4–1b). Wegen der Auswahlregel $\Delta m_S = 0$ (neben $\Delta l = \pm 1, \Delta m = 0, \pm 1$) beobachtet man die gleiche Aufspaltung der Spektrallinien wie beim normalen Zeeman-Effekt.

4.5.4.3. Anomaler Zeeman-Effekt. Hier ist von der Feinstruktur E_{njl} mit dem Drehimpulsanteil $|jlm_j\rangle$ der Eigenfunktionen auszugehen. Dieser ist zwar keine Eigenfunktion von H_m, in erster Ordnung der Störungsrechnung reicht er jedoch aus, um die Aufspaltung zu berechnen, $E_{njlm_j} = E_{njl} + g_j\mu_B Bm_j$. Der *Landé-Faktor* $g_j = 1 + [j(j+1) - l(l+1) + 3/4]/2j(j+1)$ führt zu einer unterschiedlich großen Aufspaltung einzelner Niveaus (Abb. 4.5.4–2). Deswegen ergibt sich – unter Beachtung der Auswahlregeln $\Delta j = 0, \pm 1$, $\Delta l = \pm 1$, $\Delta m_j = 0, \pm 1$ – auch eine größere Mannigfaltigkeit von Linien als beim normalen Zeeman-Effekt.

4.5.5. Stark-Effekt

Bei der Wechselwirkung eines neutralen Systems mit einem elektrischen Feld $\boldsymbol{E}$ wird die Wechselwirkungsenergie durch das Dipolmoment bestimmt, $H_e = -\check{\boldsymbol{p}}_e \cdot \boldsymbol{E}$. Da der Erwartungswert des Operators des elektrischen Dipolmomentes $\check{\boldsymbol{p}}_e$ in einem Zustand $|lm\rangle$ verschwindet, gibt es i. allg. keinen Beitrag erster Ordnung in $\boldsymbol{E}$ zur Energie. Man muß das durch $\boldsymbol{E}$ induzierte Dipolmoment $\boldsymbol{p}_e \sim \boldsymbol{E}$ berechnen und erhält – wie beobachtet – einen *quadratischen Effekt*.

Das *Wasserstoffatom* macht wegen der Entartung der Zustände gleicher Hauptquantenzahl eine Ausnahme. Die Linearkombination von $|lm\rangle$ zu verschiedenen l-Werten führt zu $\boldsymbol{p}_e \neq 0$ schon für $\boldsymbol{E} = 0$, so daß hier ein linearer *Stark-Effekt* beobachtet wird. Aus dem 2s-, 2p-Niveau entstehen z.B. im elektrischen Feld drei Niveaus mit den Energien E_2 (zweifach entartet) und $E_2 \pm 3ea_0|\boldsymbol{E}|$.

Ein homogenes elektrisches Feld hat das Potential $U = -\boldsymbol{r}\cdot\boldsymbol{E}$. Wird $-eU$ dem Coulomb-Potential $-\epsilon^2/r$ überlagert, dann bildet sich ein Potentialwall, der durchtunnelt werden kann. Dies wirkt sich dahingehend aus, daß bei starken elektrischen Feldern eine Verbreiterung der Spektrallinien beobachtet und eine Feldemission möglich wird.

4.5.6. Abschirmung einer Ladung

Eine positive Ladung in einem Metall wird durch die Leitungselektronen abgeschirmt. Um die Ladung bildet sich eine relativ begrenzte Elektronenwolke betragsmäßig gleicher Gesamtladung. Das einfachste Modell, um diese Abschirmung zu studieren, ist eine positive Ladung in einem Elektronengas (vgl. 4.10.8.). Die durch ein Potential mit der Fourier-Transformierten $v_a(q)$ induzierte *Abschirmladung* $n(r) = n_0 - \sum_q e^{iq\cdot r} v(q)\Lambda(q)$ kann aus der in erster Störungsordnung berechneten Wellenfunktion gewonnen werden. Darin bedeuten

$$\Lambda = \frac{m}{\hbar^2}\,\frac{k_F}{\pi^2} f\left(\frac{q}{2k_F}\right),\ f(x) = \frac{1}{2}\left(1 - \frac{1-x^2}{2x}\ln\left|\frac{1+x}{1-x}\right|\right).$$

Bei selbstkonsistenter Berechnung der Abschirmung muß man beachten, daß das Störpotential $v(q)$ die Summe der Potentiale von positiver Ladung $v_a(q)$ und negativer Abschirmwolke ist. Es ergibt sich zu $v(q) = v_a(q)/\varepsilon(q)$. Die

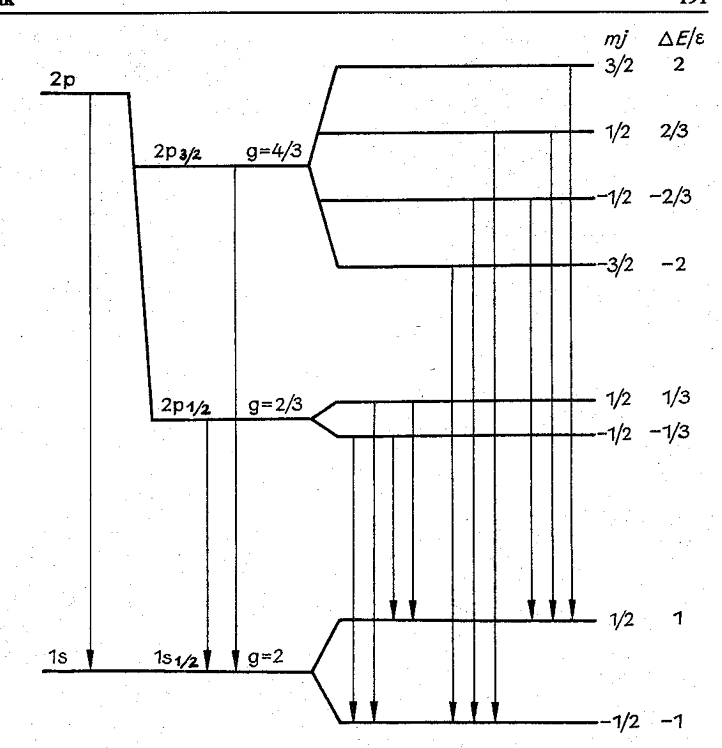


Abb. 4.5.4–2 Aufspaltung der Feinstruktur-Spektrallinien beim anomalen Zeeman-Effekt ($\varepsilon = \mu_B B$). Die Ausgangsniveaus haben keine maßstabsgerechte Lage

dielektrische Funktion $\varepsilon(q) = 1 + (4\pi \in^2/q^2)\Lambda(q)$ in dieser sogenannten Lindhardt- oder RPA-Näherung verläuft für $q \to 0$ wie $\varepsilon = 1 + (q_{TF}/q)^2$, $q_{TF} = (4/\pi)k_F/a_0$ (Thomas-Fermi-Näherung). In dieser Näherung führt ein Coulomb-Potential $v_0(r) = -\in^2/r$ zu einem abgeschirmten Potential $v(r) = -(\in^2/r)e^{-q_{TF}r} \cdot 1/q_{TF}$ charakterisiert also die Ausdehnung der Abschirmwolke. In der Lindhardt-Näherung erfaßt man in Abschirmwolke und Störpotential auch mit $\cos(2k_F r)/r^3$ oszillierende Terme, die sogenannten *Friedel-Oszillationen*.

4.6. Relativistische Quantenmechanik

So wie man aus der Energie-Impuls-Beziehung $E = H(p)$ im nichtrelativistischen Fall $H(p) = p^2/2m$ die Schrödinger-Gleichung als mögliche Form der Grundgleichung gewinnen kann, folgen aus $H(p) = c\sqrt{(mc)^2 + p^2}$ im relativistischen Fall mögliche Formen der Wellengleichung. Dieser Weg ist nicht eindeutig, für verschiedene Elementarteilchen, wie Photonen, Mesonen, Elektronen, muß man ja auch verschiedene Wellengleichungen erhalten.

4.6.1. Dirac-Gleichung

Dirac fand 1928 die nach ihm benannte, für Elektronen geltende relativistische Grundgleichung der Quantenmechanik. Eine Besonderheit der Elektronen besteht darin, daß ihre Anzahl stets erhalten bleibt, daß sie nicht – wie z. B. Photonen – erzeugt und vernichtet werden können. Das führt auf eine in $\partial/\partial t$ lineare Wellengleichung $(p - mc)\Psi(r, t) = 0$, $p = \gamma_\mu p^\mu$. Die relativistische Invarianz erfordert aber die Einführung neuer Operatoren $\gamma_\mu = (\beta, \beta\alpha)$, deren physikalische Bedeutung zu klären war. Es zeigte sich, daß die Dirac-Gleichung Teilchen mit dem *Spin* 1/2 beschreibt, den damals schon bekannten Elektronenspin also beinhaltete. Weiterhin beschreibt sie *Teilchen* und *Antiteilchen*. Dies führte Dirac zur Hypothese, daß Positronen existieren müssen. Sie wurden 1932 von Anderson erstmals experimentell nachgewiesen. Zu einem Impuls p gibt es also vier Lösungen der Dirac-Gleichung, die sich in der Spinquantenzahl $m_S = \pm 1/2$ und im Energievorzeichen unterscheiden. Eine Uminterpretation der Zustände negativer Energie führt auf die Positronen.

Die Wellenfunktionen der Dirac-Gleichung bezeichnet man als *Spinoren*. Ihre Besonderheit besteht darin, daß sie – bedingt durch den halbzahligen Spin – bei einer Drehung um 2π ihr Vorzeichen wechseln. Die Erweiterung auf die Bewegung im elektromagnetischen Feld führt auf die Dirac-Gleichung $(q - mc)\Psi = 0$, $q_\mu = p_\mu - eA_\mu$ oder $-(\hbar/i)(\partial/\partial t)\Psi = H\Psi$, $H = mc^2\beta + c\alpha \cdot (\check{p} - eA) + eU$, wenn man den Hamilton-Operator einführt. e bedeutet in diesen Gleichungen die – negative – Elektronenladung. Für das magnetische Moment des Elektrons erhält man aus der Wechselwirkungsenergie im Magnetfeld den mit dem Stern-Gerlach-Versuch in Übereinstimmung stehenden Wert mit $g = 2$.

4.6.2. Feinstrukturformel

Für das Diracsche Elektron ist im Zentralfeld $V(r)$ der Bahndrehimpuls keine Erhaltungsgröße und nicht mit H vertauschbar. Dies ist dadurch bedingt, daß das magnetische Moment des Elektrons mit dem zur Bahnbewegung ge-

hörenden Moment koppelt. Für die *Spin-Bahn-Kopplung* als relativistischen Effekt ergibt sich der in Kapitel 4.3.4. angegebene Anteil im Hamilton-Operator. Der Gesamtdrehimpuls $\boldsymbol{J} = \boldsymbol{L} + \boldsymbol{s}$ ist jedoch Erhaltungsgröße und mit H vertauschbar, ebenso auch das Quadrat des Bahndrehimpulses L^2. Die Zustände sind durch die Quantenzahlen $njlmj$ charakterisierbar, die die Eigenwerte der Operatoren H, J^2, L^2 und J_z bestimmen. Die Energieeigenwerte hängen von m_j nicht ab, beim Wasserstoffatom auch nicht von l. Die *Sommerfeldsche Feinstrukturformel*

$$E_{nj} = \frac{mc^2}{\sqrt{1 + \dfrac{\alpha}{n_r + \sqrt{k^2 - \alpha^2}}}}$$

$$= mc^2 - \frac{\epsilon^2}{2a_0}\frac{1}{n^2} - \frac{\epsilon^2}{2a_0}\frac{\alpha^2}{n^3}\left(\frac{1}{j+1/2} - \frac{3}{4n}\right) + \ldots$$

ergibt sich als Lösung der Dirac-Gleichung für das Coulomb-Potential. Sie enthält alle relativistischen Effekte, von denen die Spin-Bahn-Kopplung nur einer ist. $\alpha = \epsilon^2/(\hbar c) = 1/137{,}0360$ ist die sogenannte Sommerfeldsche Feinstrukturkonstante.

Die Entwicklung nach Potenzen von α^2 zeigt: Neben der Ruheenergie mc^2 tritt die von der nichtrelativistischen Rechnung her bekannte Bindungsenergie $-\epsilon^2/(2a_0 n^2)$ auf. Die weiteren Ordnungen sind relativistische Effekte. Neben einer Niveauverschiebung enthalten sie auch eine Aufspaltung von Niveaus gleicher Hauptquantenzahl n, je nach dem Wert des Gesamtdrehimpulses (Abb. 4.6.2–1). Diese Feinstruktur kann als verschieden große Wechselwirkungsenergie der zu Spin- und Bahnbewegung gehörenden magnetischen Momente veranschaulicht werden, da die Momente für unterschiedliche j verschieden zueinander orientiert sind.

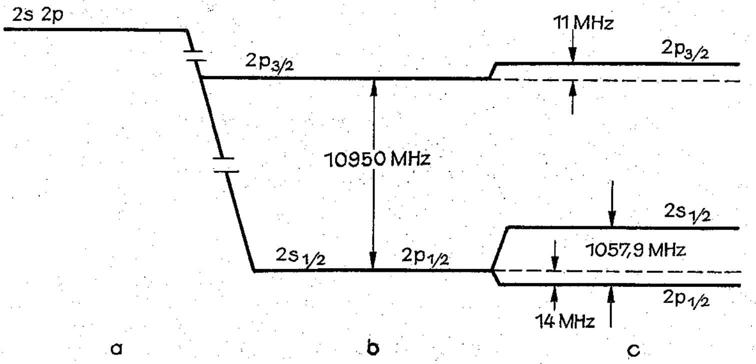


Abb. 4.6.2–1 Energieniveaus des Wasserstoffatomes als Lösung (a) der Schrödinger-Gleichung, (b) der Dirac-Gleichung (Feinstruktur) und (c) mit Strahlungskorrekturen (Lamb-Shift)

4.6.3. Quantenelektrodynamik

Die Quantenelektrodynamik ist eine Feldtheorie der elektromagnetischen Wechselwirkung. Sie beschreibt die Wechselwirkung von Leptonen (e^-, e^+, μ^-, μ^+) mit dem und über das elektromagnetische Feld. Wesentlich zu ihrer Entwicklung trug die Beobachtung von Abweichungen zu den Lösungen der Dirac-Gleichung bei. Zwei sollen hervorgehoben werden: 1. der Lamb-Shift. Lamb und Retherford beobachteten 1947 einen Energieunterschied zwischen $2s_{1/2}$- und $2p_{1/2}$-Niveau im Wasserstoffatom $E(2s_{1/2}) - E(2p_{1/2}) = h\nu$ mit $\nu = 1057{,}9$ MHz (vgl. Abb. 4.6.2–1). 2. das anomale magnetische Moment des Elektrons. Der gyromagnetische Faktor erwies sich als etwas von Zwei abweichend, $g/2 = 1{,}0011596$. Diese Größe wurde mit hoher Genauigkeit gemessen. Die Quantenelektrodynamik liefert bis zu dieser hohen Genauigkeit Resultate, die mit den beobachteten Werten in Übereinstimmung stehen. Diese Erfolge wurden bei Untersuchungen an Müonenatomen (Kern + μ^-), am Positronium ($e^+ + e^-$) und am Müonium ($\mu^+ + \mu^-$) bestätigt.

Die Grundgleichungen der Quantenelektrodynamik können mittels der Störungsrechnung ausgewertet werden, weil mit der Feinstrukturkonstanten α ein kleiner Entwicklungsparameter auftritt. Sie enthält jedoch konzeptionelle Schwierigkeiten, die sich im Auftreten von Divergenzen bemerkbar machen. Diese hängen u. a. damit zusammen, daß schon in der Elektrodynamik ein punktförmiges Elektron eine unendlich große Selbstenergie hat. Es ist zwar eine Renormierung (Beseitigung der divergenten Terme) möglich, jedoch kann die Quantenelektrodynamik kein Musterbeispiel für die Aufstellung von Feldtheorien z. B. für die starke Wechselwirkung sein.

4.7. Mehrelektronensysteme

Bei Mehrelektronensystemen treten weitere quantenmechanische Besonderheiten für Elektronen, insbesondere das Pauli-Prinzip, zutage. Sie sind schon am einfachsten System, dem Heliumatom, zu erkennen. Jedes Elektron hat vier Freiheitsgrade, seine Wellenfunktion enthält vier Koordinaten, Ortsvektor $\boldsymbol{r}$ und Spinvariable s. Um übersichtliche Formeln zu erhalten, wird teilweise die Schreibweise $\boldsymbol{r}_1, s_1 \equiv 1$; $\boldsymbol{r}_2, s_2 \equiv 2$ usw. verwendet. Die zeitunabhängige Schrödinger-Gleichung enthält den Hamilton-Operator $H(1, 2, \ldots, N)$ und ist eine partielle Differentialgleichung mit $4N$ Variablen für die Wellenfunktion $\Psi(1, 2, \ldots, N)$. Diese enthält die Information, welche Zustände von Elektronen besetzt sind, daß z. B. von den drei Elektronen des Lithiumatoms sich im Grundzustand zwei im 1s-Zustand und eins im 2s-Zustand befinden.

Der Hamilton-Operator des Heliumatoms

$$H(1,2) = \frac{\check{\boldsymbol{p}}_1^2}{2m} + \frac{\check{\boldsymbol{p}}_2^2}{2m} - \frac{2\epsilon^2}{r_1} - \frac{2\epsilon^2}{r_2} + \frac{\epsilon^2}{|\boldsymbol{r}_1 - \boldsymbol{r}_2|}$$

enthält die typischen, bei Mehrelektronensystemen auftretenden Teile: die kinetische Energie der einzelnen Elektronen, die Coulomb-Wechselwirkung der Elektronen mit dem Kern und die Coulomb-Abstoßung der Elektronen unter-

einander. Der Hamilton-Operator hat eine Symmetrieeigenschaft, die bei der quantenmechanischen Betrachtung von großer Tragweite ist. Es ist die *Symmetrie gegenüber einem Vertauschen der Teilchenkoordinaten* $H(1,2) = H(2,1)$. Daraus folgt 1. die *Ununterscheidbarkeit von Elementarteilchen* gleicher Sorte und 2. es treten zwei Typen von Eigenfunktionen auf, symmetrische und antisymmetrische:

$$H(1,2) = H(2,1) \rightarrow \begin{cases} \Psi(1,2) = \Psi(2,1) & \text{Bosonen} \\ \Psi(1,2) = -\Psi(2,1) & \text{Fermionen.} \end{cases}$$

Beide Lösungen sind in der Natur realisiert. Sie führen zu einem wesentlich verschiedenen Verhalten der Teilchen, die man *Bosonen* (Teilchen mit ganzzahligem Spin) bzw. *Fermionen* (Teilchen mit halbzahligem Spin) nennt.
Elektronen sind Fermionen. Aus der Antisymmetrie der Gesamtwellenfunktion eines Elektronensystems folgt das *Pauli-Prinzip*. Es besagt, daß ein Einteilchenzustand maximal mit einem Elektron besetzt werden kann. Für Fermionen treten nur die Besetzungszahlen 0 und 1 auf. Unter einem *Einteilchenzustand* ist dabei ein durch vier Quantenzahlen charakterisierter Zustand zu verstehen, wie er als Lösung der Schrödinger-Gleichung für ein Elektron auftritt, z. B. eine durch die Quantenzahlen $nlmm_S$ fixierte Lösung im Zentralpotential oder eine ebene Welle mit den Quantenzahlen pm_S für ein freies Elektron. Das Pauli-Prinzip findet seine glänzende experimentelle Bestätigung im *Periodischen System der Elemente*. Die Niveaus 1s, 2s, 2p, ... werden nacheinander aufgefüllt. Die Anzahl der Zustände in einer Schale bestimmt die Länge (2, 8 oder 18) der Perioden. Ohne das Pauli-Prinzip würden sich im Grundzustand alle Elektronen im 1s-Zustand befinden.

4.7.1. Heliumatom

Das Heliumatom hat zwei Elektronen, die sich im Grundzustand im 1s-Niveau (mit Spin auf bzw. Spin ab) befinden. Betrachtet man die Übergänge zwischen angeregten Zuständen, bei denen ein Elektron im 1s-Zustand verbleibt und das andere angeregt ist, dann beobachtet man das in Abb. 4.7.1–1 dargestellte Spektrum. Die Terme werden durch die Quantenzahlen des Gesamtspins S (bei zwei Elektronen Null oder Eins), des Gesamtbahndrehimpulses L (gleich dem des angeregten Elektrons), des Gesamtdrehimpulses J (er läuft von $|L-S|$ bis $L+S$) und durch die Hauptquantenzahl des Zustandes, in dem sich das angeregte Elektron befindet, klassifiziert $n^{(2S+1)}L_J$.
Zwischen den Termen mit $S=0$ (*Parahelium*) und $S=1$ (*Orthohelium*) gibt es wegen der Auswahlregel $\Delta S = 0$ keine Übergänge, der Zustand 2^3S_1 ist metastabil. Hier soll besonders auf das Fehlen des 1^3S_1-Zustandes (im Termschema des Orthoheliums) und auf den Energieunterschied zwischen 2^1S_0- und 2^3S_1-Niveau hingewiesen werden. Diese Besonderheiten können aus der Antisymmetrie der Wellenfunktion verstanden werden. Der Spinanteil der Wellenfunktion χ_{SM_S} lautet

$\chi_{00}(s_1, s_2) = [\chi_+(s_1)\chi_-(s_2) - \chi_-(s_1)\chi_+(s_2)]/\sqrt{2}$ für den Singulettzustand $S=0$ bzw.

$\chi_{10}(s_1 > s_2) = [\chi_+(s_1)\chi_-(s_2) + \chi_-(s_1)\chi_+(s_2)]/\sqrt{2}$, $\chi_{11}(s_1, s_2) = \chi_+(s_1)\chi_+(s_2)$, $\chi_{1-1}(s_1, s_2) = \chi_-(s_1)\chi_-(s_2)$

für den Triplettzustand $S=1$. Diese Funktionen sind antisymmetrisch ($S=0$) bzw. symmetrisch ($S=1$) gegen Vertauschen der Spinkoordinaten s_1, s_2. Demzufolge muß der Ortsanteil symmetrisch bzw. antisymmetrisch sein, damit die Gesamtwellenfunktion antisymmetrisch wird. Für das Parahelium im 2^1S_0-Zustand lautet sie

$$\Psi(r_1, s_1; r_2, s_2) = \frac{1}{\sqrt{2}}[\varphi_{1s}(r_1)\varphi_{2s}(r_2) + \varphi_{2s}(r_1)\varphi_{1s}(r_2)]\chi_{00}.$$

Für das Orthohelium im 2^3S_1-Zustand erhält man

$$\Psi(r_1, s_1; r_2, s_2) = \frac{1}{\sqrt{2}}[\varphi_{1s}(r_1)\varphi_{2s}(r_2) - \varphi_{2s}(r_1)\varphi_{1s}(r_2)]\chi_{1M_S}.$$

Es ist falsch, zu sagen, das Elektron 1 befindet sich im 1s-Niveau und das Elektron 2 im 2s-Niveau (oder umgekehrt). Die Variable r_1 tritt sowohl als Argument von φ_{1s} wie auch von φ_{2s} auf, ebenso r_2. Die Wellenfunktionen bringen allein zum Ausdruck, daß 1s- und 2s-Niveau mit je einem Elektron besetzt sind. So wirkt sich hier die Ununterscheidbarkeit von Elektronen aus.
Haben beide Elektronen den gleichen Ortsanteil der Wellenfunktion (z.B. beide befinden sich im 1s-Zustand), dann läßt sich nur die symmetrische Kombination der Ortsanteile bilden. Die antisymmetrische ist identisch Null, ergibt also keinen normierbaren, keinen physikalischen Zustand. Hier liest man das *Pauli-Prinzip* heraus. Wenn der Ortsanteil der beiden besetzten Zustände gleich ist (die Quantenzahlen nlm), dann muß der Spinanteil (Quantenzahl m_S) verschieden sein. Dies ist für $S=0$ (antiparallele Spins) der

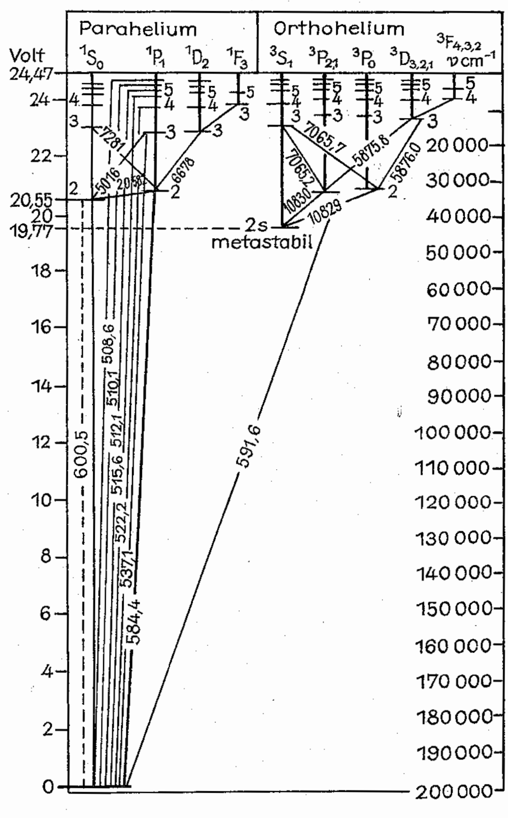


Abb. 4.7.1–1 Heliumspektrum; an den Linien sind die Wellenlängen in 10^{-10} m angegeben

Fall, nicht für $S = 1$ (parallele Spins). Damit ist das Fehlen des 1^3S_1-Zustandes im Termschema des Heliums erklärt. Bildet man mit den oben angegebenen Wellenfunktionen den Erwartungswert der Energie, dann tritt neben der kinetischen und Coulomb-Energie noch die sogenannte *Austauschenergie* (vgl. 4.7.2.) auf, $E = E_{\text{kin}} + E_{\text{Coul}} \pm E_{\text{A}}$. Je nach Symmetrie des Ortsanteiles der Wellenfunktion hat sie ein unterschiedliches Vorzeichen. Dies erklärt den Energieunterschied zwischen 2^1S_0- und 2^3S_1-Term im Heliumspektrum.

4.7.2. Hartree-Fock-Verfahren

Der Hamilton-Operator der Elektronen eines Atoms

$$H(1, 2, \ldots, N) = \sum_{n=1}^{N} \left(\frac{p_n^2}{2m} - \frac{Z \in^2}{r_n} \right) + \sum_{n<m}^{\text{Paare}} \frac{\in^2}{|r_n - r_m|}$$

enthält neben kinetischer Energie und potentieller Energie im Feld des Atomkerns die Wechselwirkung zwischen den Elektronen, in der jeder Summand von den Koordinaten zweier Elektronen abhängt. Dadurch wird die Bewegung der Elektronen miteinander gekoppelt. Die Schrödinger-Gleichung stellt eine partielle Differentialgleichung dar, die nicht separiert werden kann. Die Wellenfunktion läßt sich nicht aus Produkten von Einteilchenfunktionen aufbauen. Für solche Differentialgleichungen ist keine *analytische Lösung* auffindbar. Eine *numerische Lösung* ist wegen der zu großen Anzahl anfallender Zahlenwerte auch nicht möglich. Außerdem wären diese Zahlenwerte physikalisch im wesentlichen uninteressant. Die bei der Untersuchung von Atomen zu lösende Aufgabe besteht darin, die relativ wenigen Meßgrößen der Atome mit ausreichender Genauigkeit (gemessen an der experimentellen Genauigkeit) zu erklären bzw. näherungsweise zu berechnen. Dies erfordert, ein möglichst anschauliches Bild über den elektronischen Aufbau des Atoms zu entwickeln.

Als sehr tragfähige Konzeption hat sich die Einführung eines *mittleren Potentials* erwiesen. Dabei wird die Wechselwirkung eines herausgegriffenen Elektrons mit allen anderen durch ein mittleres Coulomb-Potential beschrieben. Die potentielle Energie des Elektrons in diesem Feld hängt nur von den Koordinaten des herausgegriffenen Elektrons ab, nicht von den Koordinaten der anderen Elektronen. In diesem Bild lassen sich Einteilchenzustände angeben, die dann von den Elektronen besetzt werden. In der Atomphysik wird durchweg mit diesem Bild gearbeitet. Bestimmte Niveaus bzw. Schalen können mit Elektronen aufgefüllt werden. Dies zeigt, daß die genannte Konzeption dem experimentell beobachteten Material sehr gut entspricht.

Das *Hartree-Fock-Verfahren* ist die mathematische Realisierung dieser Konzeption. Die Gesamtwellenfunktion wird aus Produkten von Einteilchenfunktionen $\psi_k(\boldsymbol{r}, s) = \varphi_{nlm}(\boldsymbol{r})\chi_{m_S}(s)$ aufgebaut. Dabei werden Linearkombinationen derart gewählt, daß die Gesamtwellenfunktion antisymmetrisch ist. Dies ist mit der *Slater-Determinante*

$$\Psi(1, 2, \ldots, N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \psi_{k_1}(1) & \psi_{k_1}(2) & \ldots & \psi_{k_1}(N) \\ \psi_{k_2}(1) & \psi_{k_2}(2) & \ldots & \psi_{k_2}(N) \\ \vdots & \vdots & & \vdots \\ \psi_{k_N}(1) & \psi_{k_N}(2) & \ldots & \psi_{k_N}(N) \end{vmatrix}$$

erfüllt. Das Vertauschen zweier Koordinaten entspricht dem Vertauschen zweier Spalten in der Determinante. Das ist aber einem Vorzeichenwechsel äquivalent. Die Funktionen ψ_k werden so bestimmt, daß der Erwartungswert der Energie minimal wird. Dann ist die Slater-Determinante die beste Näherung der exakten Wellenfunktion. Daraus folgen die *Hartree-Fock-Gleichungen*

$$h\psi_k = E_k\psi_k, \quad h = \frac{\hat{p}^2}{2m} + v_c + v_{xc}.$$

Man führt den quantenmechanischen Ausdruck für die Teilchendichte $n(\boldsymbol{r}) = 2 \sum^{\text{bes.}} |\varphi_{nlm}(\boldsymbol{r})|^2$ als Summe über die Wahrscheinlichkeitsdichten der besetzten Zustände ein. Dabei ist jeder Ortszustand mit zwei Elektronen besetzt; daher rührt der Faktor 2. $v_c = -\frac{Z\in^2}{r} + \int d\tau' \frac{\in^2}{|\boldsymbol{r} - \boldsymbol{r}'|} n(\boldsymbol{r}')$ enthält neben dem Coulomb-Potential des Atomkerns das *Coulomb-Potential* der Ladungsverteilung $en(\boldsymbol{r})$. Dieser Anteil stellt das mittlere Coulomb-Potential dar, das anstelle der paarweisen Coulomb-Wechselwirkung auftritt. Das *Austauschpotential* v_x – bei Erweiterung um Korrelationsbeiträge auch Austausch- und Korrelationspotential v_{xc} genannt – ist ein Integraloperator. Oft ist die *lokale-Dichte-Näherung* $v_{xc} = -\alpha 3/2(3/\pi)^{1/3} \in^2 n^{1/3}$ ausreichend. Der Faktor α schwankt je nach betrachteter Substanz und nach Art der Rechnung zwischen den Grenzen $\alpha = 1$ (Slater-Austauschpotential) und $\alpha = 2/3$ (Kohn-Sham-Gaspar-Austauschpotential). Bei Elektronen mit parallelem Spin ist der zugehörige Ortsanteil der Wellenfunktion antisymmetrisch gegen Vertauschung der Ortskoordinaten (vgl. Orthohelium). Daher ist die Wahrscheinlichkeit, diese beiden Elektronen an eng benachbarten Orten anzutreffen, klein, für gleiche Orte ist sie Null. Aufgrund der Antisymmetrie der Gesamtwellenfunktion kommt die Coulomb-Abstoßung zwischen Elektronen parallelen Spins nicht voll zur Wirkung. Dieser Effekt wird im (negativen) Austauschpotential erfaßt.

Die Hartree-Fock-Gleichungen bilden den Ausgangspunkt für die überwiegende Zahl der Rechnungen zur Elektronenstruktur der Atome, Moleküle und Festkörper. Sie können zwanglos durch Spinbahnkopplung (bei hoher Ordnungszahl erforderlich) ergänzt werden. Man muß beachten, daß eine Slater-Determinante keine Eigenfunktion zu J^2 ist. Um die Multiplettstruktur von Atomspektren zu beschreiben, muß man daher von Linearkombinationen von Slater-Determinanten ausgehen. Bei der Berechnung angeregter Zustände muß man mit einer anderen Teilchendichte $n(\boldsymbol{r})$ arbeiten als für den Grundzustand, da andere Zustände besetzt sind, sonst erhält man falsche Anregungsenergien.

4.7.3. Chemische Bindung

Die Quantenmechanik eröffnete auch den Weg zum Verständnis der chemischen Bindung (vgl. 5.7.). Da die Schrödinger-Gleichung bei Mehrelektronenproblemen nicht analytisch lösbar ist, kommt es bei genäherter Berechnung darauf an, in dem Ansatz für die Wellenfunktion möglichst viele physikalische Vorstellungen über die Elektronenstruktur der Moleküle einfließen zu lassen. Diese Vorstellungen konnten mit wachsender Erfahrung verfeinert werden. Dabei kam dem genauen Studium der einfachsten Systeme besondere Bedeutung zu.

4.7.3.1. Wasserstoffmolekülion H_2^+. Es ist das einfachste System mit nur einem Elektron. Es gelang, die zugehörige Schrödinger-Gleichung analytisch zu lösen. Dadurch konnten für kompliziertere Systeme entwickelte Näherungsme-

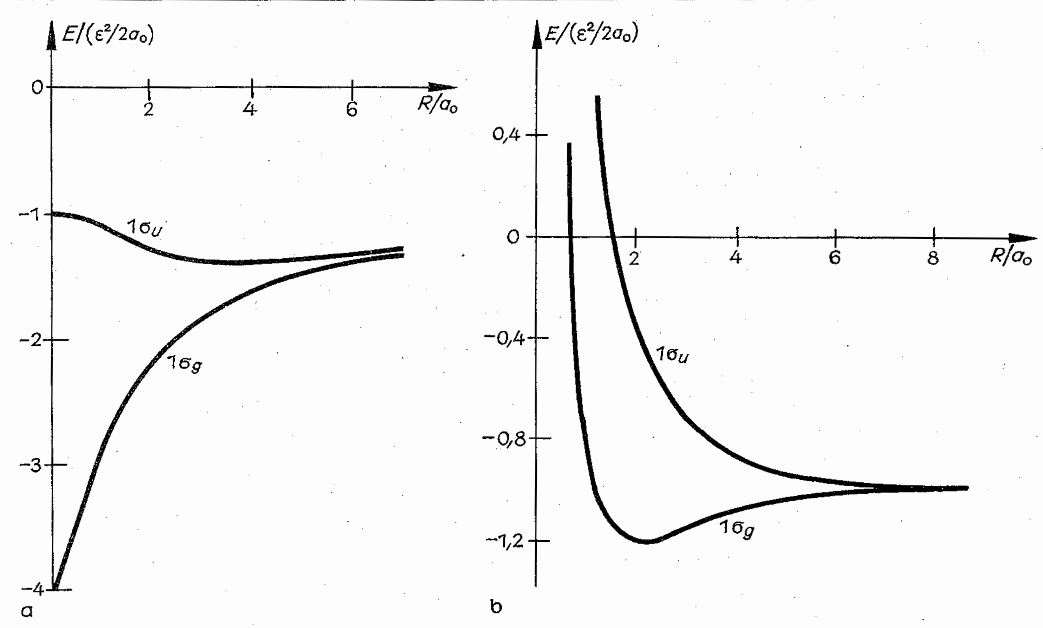


Abb. 4.7.3–1 Energie des H_2^+-Ions (a) ohne und (b) mit Kernabstoßung

thoden getestet werden. Legt man die z-Achse in Richtung der Verbindungslinie der beiden Kerne, dann ist die z-Komponente des Bahndrehimpulses eine Erhaltungsgröße, da das System rotationsinvariant ist. Die Zustände zu $m = 0$ bzw. ± 1 bzw. ± 2 bezeichnet man als σ- bzw. π- bzw. δ-Zustände. Weiterhin hat das Molekül ein Inversionszentrum in der Mitte der Verbindungslinie der Kerne. Daher verhalten sich die Lösungen bezüglich einer Inversion gerade (g) bzw. ungerade (u). Abb. 4.7.3–1 zeigt die Energien der Elektronenzustände in der durch die beiden Kerne gebildeten Potentialdoppelmulde als Funktion des Kernabstandes R. Für $R \to 0$ ergeben sich die Energien des He^+-Ions. Um die *Bindungs*- bzw. *Dissoziationsenergie* zu berechnen, muß die Abstoßung zwischen den Protonen berücksichtigt werden. Nur der $1\,\sigma_g$-Zustand führt zu einem gebundenen Zustand. Der Kernabstand beträgt $R_0 \approx 2a_0$ und die Dissoziationsenergie $E_D \approx 2{,}7$ eV. Die Wellenfunktionen, auch *molekulare Orbitale* (MO) genannt, der in der Abb. dargestellten Zustände können als gerade $\varphi_g = \varphi_{1s}^a + \varphi_{1s}^b$ bzw. ungerade $\varphi_u = \varphi_{1s}^a - \varphi_{1s}^b$ Kombinationen von 1s-Funktionen beider Kerne (a bzw. b) verstanden werden. Mit der Annäherung der Kerne spaltet der atomare 1s-Zustand auf, wobei der symmetrische Zustand, in dem die Aufenthaltswahrscheinlichkeit zwischen den beiden Kernen erhöht ist, energetisch günstiger liegt.

4.7.3.2. Wasserstoffmolekül H_2. Von ihm sollen nur die Zustände mit dem Gesamtspin $S = 0$ diskutiert werden, da unter ihnen der Grundzustand (Kernabstand $R_0 \approx 1{,}4\,a_0$, Bindungsenergie $E_{H_2} = 0{,}347$ Ryd) zu finden ist. Sie haben einen gegen Vertauschen der Koordinaten symmetrischen Ortsanteil wie das Parahelium. Aus den MO des H_2^+ lassen sich die drei Funktionen $\Phi_{g+} = \varphi_g(r_1)\varphi_g(r_2)$, $\Phi_{g-} = \varphi_u(r_1)\varphi_u(r_2)$ und $\Phi_u = \varphi_g(r_1)\varphi_u(r_2) + \varphi_u(r_1)\varphi_g(r_2)$ bilden. Sie enthalten alle drei Beiträge vom Typ $\varphi^a(r_1)\varphi^a(r_2)$ bzw. $\varphi^b(r_1)\varphi^b(r_2)$. Diese bestimmen die Wahrscheinlichkeit, daß sich beide Elektronen am gleichen Proton aufhalten. Da diese Situation nicht der physikalischen Vorstellung vom Wasserstoffmolekül entspricht, gingen HEITLER und LONDON von dem Ansatz $\Phi^{HL} = \varphi_{1s}^a(r_1)\varphi_{1s}^b(r_2) + \varphi_{1s}^b(r_1)\varphi_{1s}^a(r_2)$ aus, der einer Linearkombination molekularer Orbitale $\Phi_{g+} - \Phi_{g-}$ entspricht, in der Produkte von Einteilchenfunktionen zum gleichen Kernort nicht mehr auftreten. Dies führte schon zu einer guten Beschreibung des H_2-Moleküls. Verfeinerte Rechnungen mit $\Phi = \alpha(\Phi_{g+} - \Phi_{g-}) + \beta(\Phi_{g+} + \Phi_{g-})$ führten auf $\beta/\alpha = 0{,}24$. Dieser Wert liegt weitaus näher am Heitler-London-Fall ($\beta/\alpha = 0$) als am Fall eines reinen molekularen Orbitals ($\beta/\alpha = 1$).

Die Bindung ergibt sich durch eine Erhöhung der Ladungsdichte in dem Bereich, in dem sich die atomaren Potentialtöpfe überlappen, so daß die Elektronen an beiden Kernen gebunden sind. Man spricht von einer *homöopolaren Bindung*. Sie beinhaltet eine *Sättigung* der Bindung. Im Falle des Wasserstoffes heißt das, es gibt kein H_3-Molekül. Sind unter den Valenzelektronen auch p-Elektronen, wie bei Kohlenstoff, Stickstoff oder Sauerstoff, dann sind Bindungen mit größerer Mannigfaltigkeit möglich. Die Sättigung hängt damit zuammen, daß sich – beim H_3-Molekül – aus den 1s-Zuständen des Wasserstoffes keine Gesamtwellenfunktion konstruieren läßt, die zwischen allen drei Protonenpaaren eine erhöhte Ladungsdichte hat.

4.7.3.3. Metallische Bindungen ergeben sich in Systemen, in denen die molekularen Orbitale energetisch bevorzugt sind. Sie baut also auf Einelektronenzustände auf, die über das gesamte System ausgebreitet sind. Die Bloch-Funktionen des Festkörpers sind ein Beispiel dafür. Die metallische Bindung zeigt keine Sättigung. Beim Hinzufügen eines weiteren Atoms werden die molekularen Orbitale nur weiter ausgedehnt.

4.7.3.4. Ionische Bindungen zwischen verschiedenen Atomen, z. B. zwischen Na und Cl, sind dadurch gekennzeichnet, daß ein Elektron des einen Atoms (Na) zum anderen Atomplatz wechselt. Chlor kann ein negativ geladenes Ion

Cl^- bilden. Die Bindungsenergie des überschüssigen Elektrons beträgt 3,71 eV. Die Ionisationsenergie des Natriums ist 5,14 eV. Die Differenz beider Energien wird übertroffen von der Bindungsenergie $Na^+ + Cl^- \rightarrow NaCl + 4{,}51$ eV. Je Molekül NaCl wird also eine Energie von 3,08 eV frei.

4.8. Grundlagen der Statistik

Das makroskopische Verhalten eines Systems, z. B. eines Gases, wird im thermodynamischen Gleichgewicht durch wenige Parameter, wie Volumen und Temperatur, gekennzeichnet. Aus mikroskopischer Sicht, d. h. in den Positionen und Geschwindigkeiten der einzelnen Moleküle, bestehen jedoch wesentliche Unterschiede, sowohl beim Vergleich des Zustandes eines Systems zu verschiedenen Zeitpunkten wie auch beim Vergleich verschiedener Systeme mit gleichen makroskopischen Parametern.
Die Aufgabe der Statistik besteht darin, den Zusammenhang zwischen der Bewegung der Bestandteile des Systems entsprechend den Grundgleichungen der Mechanik oder Quantenmechanik und den thermodynamischen Eigenschaften des Systems herzustellen. Dabei zeigt sich, daß Systeme mit großer Teilchenzahl Besonderheiten aufweisen, die allein durch die große Teilchenzahl bedingt sind. Die Statistik ist weiterhin in der Lage, die Größe der Schwankungen um den thermodynamischen Gleichgewichtszustand zu erfassen. Da die Statistik einen Zusammenhang zwischen den thermodynamischen Eigenschaften und den Bewegungsgleichungen herstellt, ist es möglich, die Materialkonstanten (spezifische Wärme, Kompressibilität usw.) auf Eigenschaften der Teilchen des Systems bzw. der zwischen diesen wirkenden Kräfte zurückzuführen.
Als zentrales Bindeglied kann man den Zusammenhang zwischen Entropie und Verteilung der Mikrozustände ansehen. Er liefert aber nicht nur die Beziehung zwischen thermodynamischem Potential und Teilcheneigenschaften im thermodynamischen Gleichgewichtszustand, sondern stellt eine Erweiterung des Entropiebegriffes auf beliebige Nichtgleichgewichtszustände dar.

4.8.1. Gesamtheiten

Unter einem Zustand ν ist in der Statistik ein Mikrozustand zu verstehen, der in der klassischen Statistik (vgl. 4.9.) durch die Angabe der Orte und Impulse $\boldsymbol{r}_i$, $\boldsymbol{p}_i$ aller Teilchen bestimmt ist, in der Quantenstatistik (vgl. 4.10.) durch ein Spektrum von Eigenwerten und zugehörigen Besetzungszahlen.

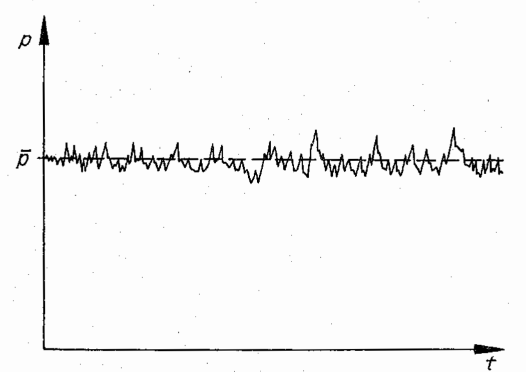


Abb. 4.8.1–1 Fluktuationen des Druckes

Zu vorgegebenen makroskopischen Parametern gibt es viele *mögliche Zustände*. Wird z.B. ein ideales Gas in einem abgeschlossenen System betrachtet, haben Teilchenzahl N, Volumen V und innere Energie U feste Werte. Die Energie eines Zustandes $E_\nu = \sum_i p_i^2/2m$ ist in diesem Fall gleich der Summe der kinetischen Energien aller Teilchen. Es gibt aber viele verschiedene Zustände ν mit $E_\nu = U$; bis auf die Bedingung $\sum_i p_i^2/2m = U$ können die Einzelimpulse $\boldsymbol{p}_i$ willkürlich gewählt werden. Für die Positionen $\boldsymbol{r}_i$ gibt es überhaupt keine Einschränkungen. Die Vielfalt der möglichen Zustände, die mit bestimmten makroskopischen Parametern verträglich sind, wird als eine *Gesamtheit* bezeichnet. Je nach dem, welche Parameter vorgegeben sind, werden verschiedene Gesamtheiten betrachtet.
Gesamtheiten werden durch Angabe der *Wahrscheinlichkeiten* w_ν charakterisiert, dafür, daß der Zustand ν in der Gesamtheit enthalten ist. Die Summe der Wahrscheinlichkeiten über alle Zustände ist auf Eins normiert, $\sum_\nu w_\nu = 1$. Ist ein Zustand ν mit den vorgegebenen makroskopischen Parametern nicht vereinbar, dann ist er in der Gesamtheit nicht enthalten, es ist $w_\nu = 0$. Die Grundannahme der Statistik besteht darin, daß man allen mit den Parameterwerten verträglichen Zuständen des abgeschlossenen Systems die gleiche Wahrscheinlichkeit zuordnet. Man wählt die *a-priori-Wahrscheinlichkeit* für alle möglichen Zustände gleich groß. Dieses Postulat kann nicht bewiesen werden, es wird dadurch gestützt, daß alle aus der Statistik gewonnenen Aussagen in Übereinstimmung mit den experimentellen Erfahrungen stehen.
Statt der Zeitmittelwerte berechnet man in der Statistik Mittelwerte über mögliche Zustände, sogenannte *Scharmittelwerte*, wobei als Schar die zu den gegebenen Bedingungen gehörige Gesamtheit gewählt wird. Zur Begründung dieses Vorgehens wird die *Ergodenhypothese* herangezogen. Sie besagt, daß ein System im Zeitablauf jedem der in der Schar betrachteten Zustände beliebig nahekommt.
4.8.1.1. Mittelwerte. Ein System durchläuft im Zeitablauf verschiedene mögliche Zustände. Zum Beispiel ändern sich in einem Gas die Impulse der einzelnen Teilchen durch Stöße untereinander und mit der Wand. Durch die Bewegung verschieben sich laufend Positionen. Makroskopische Meßwerte stellen *Zeitmittelwerte* dar. So wird der Druck, die Kraft je Fläche, durch die Impulsänderung der an der Fläche reflektierten Teilchen je Zeit bestimmt. Der Druck ist deswegen zeitlich konstant, weil während der Meßzeit viele Stöße auf die Fläche stattfinden und der Mittelwert der Impulsübertragung – gemittelt über die Meßzeit – zeitunabhängig ist (Abb. 4.8.1–1). Könnte die Meßzeit auf etwa die Stoßzeit verkürzt werden, dann würde man beträchtliche Fluktuationen feststellen.
4.8.1.2. Meßwerte von physikalischen Größen werden in der Statistik berechnet, indem man den Wert einer physikalischen Größe A im Zustand ν, d. h. A_ν, mit der Wahrscheinlichkeit des Auftretens w_ν dieses Zustandes multipliziert und über alle Zustände ν summiert

$$\bar{A} = \sum_\nu w_\nu A_\nu, \quad \sum_\nu w_\nu = 1.$$

$\bar{A}$ wird auch als *Erwartungswert* bezeichnet, z. B. ist $U = \sum_\nu w_\nu E_\nu$. Die *mittlere quadratische Schwankung*

$$(\Delta A)^2 = \sum_\nu w_\nu (A_\nu - \bar{A})^2 = \overline{A^2} - \bar{A}^2$$

charakterisiert die Fluktuationen der Werte A_ν um den Erwartungswert. Für beliebige Gesamtheiten, charakterisiert durch einen Satz von Wahrscheinlichkeiten w_ν, $\sum_\nu w_\nu = 1$, bezeichnet man als Entropie die Größe

$$S = -k \sum_\nu w_\nu \ln w_\nu .$$

Diese Form wurde z. B. durch den *H-Theorem* von BOLTZMANN (vgl. 4.9.6.) nahegelegt. Diese Definition bleibt über die Gleichgewichtsthermodynamik hinaus gültig. Sie gilt auch für Nichtgleichgewichtszustände und hat sich ebenso in der Informationstheorie bewährt. S hat die Eigenschaften der in der Thermodynamik eingeführten Entropie. Sie ist positiv definit und eine extensive, additive Größe. Sie ist ein Maß für die Unordnung. Maximale Bestimmtheit liegt dann vor, wenn sich die möglichen Zustände auf einen einzigen reduzieren. Wird diesem den Wert $\nu = 1$ zugeordnet, dann gilt in diesem Falle $w_1 = 1$ und $w_\nu = 0$ für $\nu \neq 1$. Wegen $\ln 1 = 0$ und $\lim_{x\to 0} x \ln x = 0$ hat S im Zustand maximaler Bestimmtheit seinen kleinstmöglichen Wert $S = 0$. S erreicht seinen Maximalwert, wenn alle w_ν gleich groß sind (Abb. 4.8.1-2). Das ist der Zustand größter Unordnung.

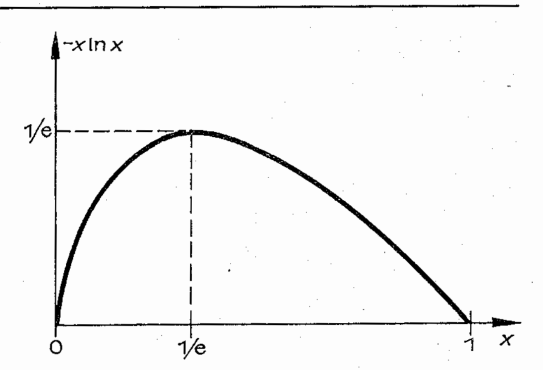


Abb. 4.8.1-2 Form der Funktion $-x \ln x$

4.8.2. Thermodynamische Gesamtheiten

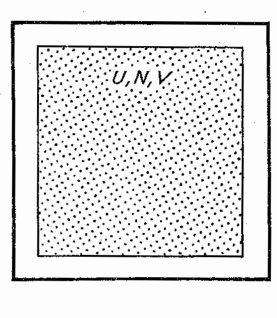
Eine Gesamtheit ist durch die Angabe der w_ν bestimmt. w_ν gibt an, mit welcher Wahrscheinlichkeit der Zustand ν in der Gesamtheit enthalten ist. Zur Berechnung thermodynamischer Eigenschaften interessieren insbesondere solche Gesamtheiten, die den thermodynamischen Gleichgewichtszustand charakterisieren. Die wichtigsten Beispiele sind die mikrokanonische, die kanonische und die großkanonische Gesamtheit. Sie unterscheiden sich in den aufgeprägten Parametern (Abb. 4.8.2-1). Ein abgeschlossenes System, für das Energie U, Teilchenzahl N und Volumen V vorgegeben sind, wird durch eine *mikrokanonische Gesamtheit* charakterisiert. Steht das System in Kontakt mit einem Wärmebad, so daß Temperatur bzw. mittlere Energie aufgeprägt werden, dann liegt eine *kanonische Gesamtheit* vor. Kann mit dem Reservoir auch ein Teilchenaustausch stattfinden, so daß neben der Temperatur die mittlere Teilchenzahl bzw. das chemische Potential vorgegeben sind, dann ist eine *großkanonische Gesamtheit* zu betrachten.

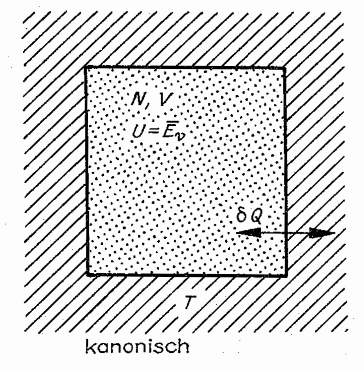
4.8.2.1. Mikrokanonische Gesamtheit. Sie umfaßt Systeme in Zuständen ν, deren Energie E_ν in einem vorgegebenen Bereich $U \ldots U + \Delta U$ liegt. Alle möglichen Zustände bekommen entsprechend der grundlegenden Annahme der Statistik die gleiche Wahrscheinlichkeit zugeordnet, so daß w_ν durch

$$w_\nu = \begin{cases} \text{konst.} & \text{für } U \leqq E_\nu \leqq U + \Delta U \\ 0 & \text{sonst} \end{cases}$$

gegeben sind. Betrachtet man z. B. ein ideales Gas, dann ist $E_\nu = \sum_i p_i^2/2m$. Zur Summe über ν tragen also nur solche Zustände bei, für die die Summe der Impulsquadrate einen vorgegebenen Wert aufweist. Das sind sehr viele Zustände. Schon bei einem Teilchen gilt $p^2 = \text{konst.}$ für alle Impulse auf einer Kugeloberfläche. Bei zwei Teilchen muß man alle Aufteilungen $E_\nu = \varepsilon_1 + \varepsilon_2$ betrachten, wobei ε_1 und ε_2 die Energien der beiden Teilchen bedeuten. Zu festen Werten ε_1 bzw. ε_2 gehören wieder alle Impulse p_1 bzw. p_2 auf Kugeloberflächen mit durch p_1 bzw. p_2 bestimmten Radien. Mit wachsender Teilchenzahl wächst die Anzahl der Realisierungsmöglichkeiten einer Energie E und damit die Zahl der zu der Gesamtheit gehörenden Zustände wie $D(E)\Delta E \sim E^{(3N-2)/2}\Delta E$, also außerordentlich rasch, an.

4.8.2.2. Kanonische Gesamtheit. Bei ihr kann das betrachtete System mit einem Wärmereservoir Energie austauschen. Ein *Wärmereservoir* ist ein großes System, dessen Eigenschaften durch den Kontakt mit dem von uns betrachteten System nicht in meßbarer Weise beeinflußt





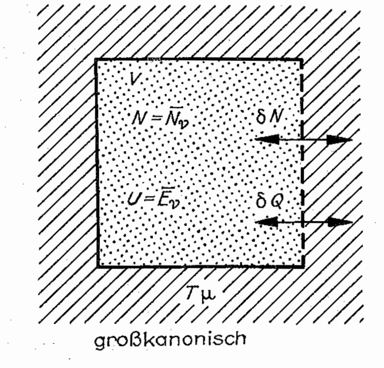


Abb. 4.8.2-1 Thermodynamische Gesamtheiten

werden soll. Das Gesamtsystem – Reservoir plus betrachtetes Teilsystem – wird als abgeschlossen angenommen. Die Gesamtenergie E_{ges} ist dann vorgegeben. Alle möglichen Zustände des Gesamtsystems bilden eine mikrokanonische Gesamtheit, sie haben die gleiche Wahrscheinlichkeit. Zu jedem dieser möglichen Zustände gehört eine Aufteilung $E_{ges} = E_R + E_\nu$ der Gesamtenergie in die des Reservoirs und die des Teilsystems. Eine Aufteilung mit einem bestimmten Wert E_ν für die Energie des Teilsystems tritt mit einer Wahrscheinlichkeit w_ν auf. Dessen innere Energie ist daher durch $U = \sum_\nu w_\nu E_\nu$ gegeben. w_ν bzw. U wird durch Eigenschaften des Wärmereservoirs bestimmt. Der aufgeprägte Parameter wird in der Thermodynamik als Temperatur T bezeichnet.

Die Forderung, daß die Entropie $S = -k \sum_\nu w_\nu \ln w_\nu$ für gegebene mittlere Energie U ein Maximum annimmt, führt auf $w_\nu = e^{-E_\nu/kT} / \left(\sum_{\nu'} e^{-E_{\nu'}/kT}\right)$ für eine kanonische Gesamtheit. Die Summe über ν läuft über alle Energien E_ν und in jedem Energiebereich ΔE über die in einer mikrokanonischen Gesamtheit enthaltenen möglichen Zustände. w_ν ist am größten für $E_\nu = 0$. Mit größter Wahrscheinlichkeit hat das System jedoch eine Energie nahe U, da die Zahl der Zustände ν gleicher Energie mit steigender Energie stark anwächst. Man schreibt die Summe $\sum_\nu w_\nu = \sum_E w_E D(E) \Delta E$ in eine Summe über die Energie um, indem man die *Zustandsdichte* $D(E)$ einführt. $D(E)\Delta E$ bedeutet die Zahl der Zustände mit einer Energie zwischen E und $E + \Delta E$. Es gilt $D(E) \sim E^{(3N-2)/2}$, so daß bei großer Teilchenzahl $N \approx 10^{23} D(E)$ außerordentlich rasch mit der Energie anwächst. Wegen des exponentiellen Abfallens von w_E hat das Produkt $w_E D(E)$ in der Umgebung von $E = U$ ein scharfes Maximum (Abb. 4.8.2–2), so daß zur Summe über E nur ein schmaler Bereich um U beiträgt. Die relative Breite dieses Bereiches ist zu $1/\sqrt{N}$ proportional. Diese Größenordnung hat daher auch Schwankungserscheinungen.

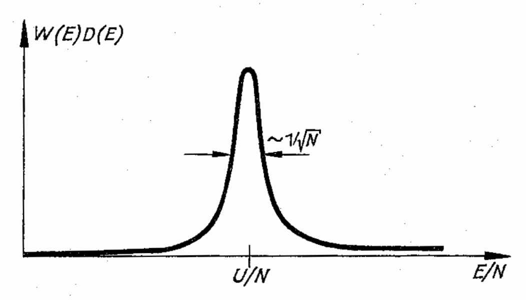


Abb. 4.8.2–2 $w(E)D(E)$ bei kanonischer Gesamtheit

4.8.2.3. Großkanonische Gesamtheit. Bei ihr kann in Erweiterung zur kanonischen Gesamtheit auch die Teilchenzahl des betrachteten Teilsystems variieren, da ein Teilchenaustausch mit dem Reservoir möglich sein soll. Die mittlere Teilchenzahl N wird – neben der mittleren Energie U – aufgeprägt. Im thermodynamischen Gleichgewicht wird sie – vom Standpunkt der Thermodynamik aus gesehen – dadurch bestimmt, daß die chemischen Potentiale von Reservoir und Teilsystem gleich sind. Da die Eigenschaften des Reservoirs und damit auch sein chemisches Potential μ durch den Kontakt mit dem Teilsystem nicht beeinflußt werden sollen, kann man schließen: Der aufgeprägte Parameter ist die Größe, die in der Thermodynamik als *chemisches Potential* bezeichnet wird. Aus der Forderung, daß die Entropie $S = -k \sum_\nu w_\nu \ln w_\nu$ für gegebene mittlere Energie $U = \sum_\nu w_\nu E_\nu$ und gegebene mittlere Teilchenzahl $N = \sum_\nu w_\nu N_\nu$ ein Maximum annehmen soll, folgt

$$w_\nu = e^{-(E_\nu - \mu N_\nu)/kT} / \left(\sum_{\nu'} e^{-(E_{\nu'} - \mu N_{\nu'})/kT}\right)$$

für eine großkanonische Gesamtheit. Die Summe über ν umfaßt hier die Summe über alle Teilchenzahlen N_ν und für einen festen Wert N_ν über alle bei der kanonischen Gesamtheit genannten Zustände.

Die *Zustandssumme*

$$Z_{kan} = \sum_\nu e^{-\beta E_\nu}, \quad Z_{großkan} = \sum_\nu e^{-\beta(E_\nu - \mu N_\nu)},$$

$\beta = 1/(kT)$, ist eine praktische Hilfsgröße, die die Berechnung thermodynamischer Größen in der Statistik erleichtert. Aus Z folgen durch Differentiation z. B. $w_\nu = -(1/\beta)\,(\partial \ln Z/\partial E_\nu)$ und $U = -(\partial \ln Z/\partial \beta)$.

4.8.3. Thermodynamische Potentiale

Durch die Beziehung $S = -k \sum_\nu w_\nu \ln w_\nu$ für die Entropie wurde der Zusammenhang zwischen Statistik und Thermodynamik hergestellt. Sie liefert das thermodynamische Potential $S(U, V, N)$ (vgl. 2.6.4.). Die freie Energie $F(T, V, N) = U - TS$ ist besonders einfach zu gewinnen, wenn man die Zustandssumme Z kennt. Man erhält

$$F(T, V, N) = -kT \ln Z$$

für eine kanonische Gesamtheit und $F(T, V, N) = -kT \ln Z + \mu N$ für eine großkanonische Gesamtheit, wobei im letzteren Falle N die mittlere Teilchenzahl bedeutet.

4.8.4. Das Auftreten großer Zahlen

Die in der Thermodynamik betrachteten Systeme haben größenordnungsmäßig die Menge 1 Mol bzw. die Teilchenzahl 10^{23}. Die als Wärmereservoir fungierenden Systeme sollen im Vergleich dazu groß sein. Beim Rechnen mit solchen großen Teilchenzahlen treten Besonderheiten auf, die an einem Beispiel erläutert werden sollen.

N_{ges} Teilchen sollen in einem Volumen $V_{ges} = V_R + V$ verteilt werden, gefragt wird nach der Wahrscheinlichkeit w_n, n Teilchen im Teilvolumen V vorzufinden. Ist kein Teil des Gesamtvolumens ausgezeichnet, so ist die Wahrscheinlichkeit p, daß ein Teilchen in das Volumen V zu liegen kommt, $p = V/V_{ges}$. Die Wahrscheinlichkeit, daß das Teilchen in V_R zu liegen kommt, ist $V_R/V_{ges} = 1 - p$. Bei homogener Verteilung der Teilchen in V_{ges} ist die Teilchendichte konstant und gleich N_{ges}/V_{ges}. In diesem Fall ist die Zahl der Teilchen im Volumen V durch $N = (N_{ges}/V_{ges})V = pN_{ges}$ gegeben.

Die Wahrscheinlichkeit, ganz bestimmte n Teilchen in V und die restlichen $N_{ges} - n$ Teilchen in V_R vorzufinden, ist gleich dem Produkt der Wahrscheinlichkeiten $p^n(1-p)^{N_{ges}-n}$, n Faktoren p und $N_{ges} - n$ Faktoren $(1-p)$. Da die Anzahl der Möglichkeiten, n Teilchen aus N_{ges} aus-

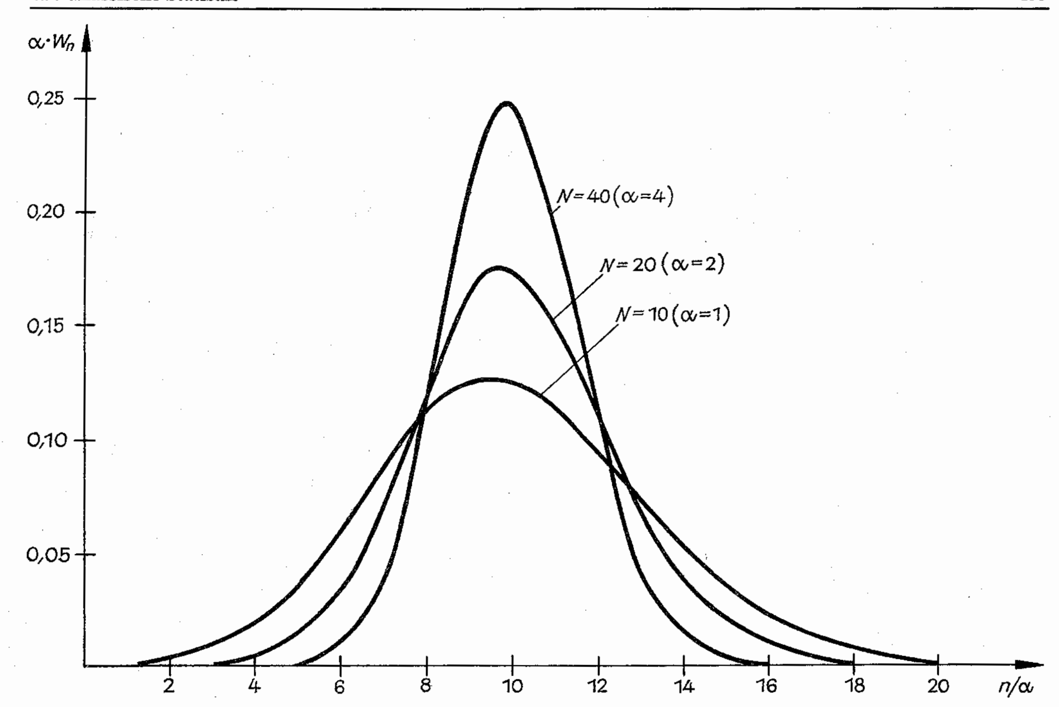


Abb. 4.8.4–1 Poisson-Verteilung für $N = 10, 20, 40$. In allen drei Fällen ist die Fläche auf Eins normiert. Der Maßstabsfaktor α wurde so gewählt, daß die Maxima beim gleichen Wert n/α liegen

zuwählen durch $\binom{N_{\text{ges}}}{n}$ gegeben ist, erhält man $w_n = \binom{N_{\text{ges}}}{n} p^n (1-p)^{N_{\text{ges}}-n}$.

Für den Fall, daß das betrachtete Volumen V klein gegen das Volumen V_R des Reservoirs ist, d.h. im Grenzfall $p \to 0$, vereinfacht sich w_n zur sogenannten *Poisson-Verteilung* $w_n = N! \mathrm{e}^{-N}/n!$. Für den *Mittelwert* der Anzahl der Teilchen im Volumen V ergibt sich daraus $\bar{n} = \sum_n n w_n = N$, wie man es erwartet. Mit dem Mittelwert $\overline{n^2} = \sum_n n^2 w_n = N^2 + N$ kann man die *mittlere quadratische Schwankung* der Teilchenzahl $(\Delta n)^2 = \overline{(n-N)^2} = \overline{n^2} - \bar{n}^2 = N$ berechnen. Die *relative Schwankung*

$$\Delta n/N = N^{-1/2}$$

ist bei großer Teilchenzahl außerordentlich klein. Bei $N = 10^{24}$ beträgt sie $\Delta n/N = 10^{-12}$. Von dieser Größenordnung sind die in thermodynamischen Systemen auftretenden Schwankungen. Die Poisson-Verteilung ist also nur in einer sehr engen Umgebung um $n = N$ wesentlich von Null verschieden (Abb. 4.8.4–1). Die Kurvenform von w_n in diesem Bereich, also für $|n - N| \ll N$,

$$w_n = [1/\Delta n\sqrt{2\pi}]\exp\left[-\frac{1}{2}\left(\frac{n-N}{\Delta n}\right)^2\right]$$

ist eine Gauß-Verteilung.

4.9. Klassische Statistik

Betrachtet wird ein System mit N Teilchen und dem Volumen V. Es stehe mit einem Wärmebad der Temperatur T in Kontakt, so daß sein Gleichgewichtszustand in der Statistik durch eine kanonische Gesamtheit beschrieben wird, bei der ein Zustand ν mit der Wahrscheinlichkeit $w_\nu \sim \mathrm{e}^{-\beta E_\nu}$ ($\beta = 1/kT$) besetzt ist. Das System mit N Teilchen hat $3N$ Freiheitsgrade. Ein *Zustand* ν *eines klassischen Systems* ist durch die Angabe der Orte $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N$ und der Impulse $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N$ aller Teilchen bestimmt. Die Energie des Zustandes $E_\nu = H(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N; \boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ ist durch den Wert der Hamilton-Funktion gegeben (vgl. 2.1.8.2.). H ist die Summe von kinetischer und potentieller Energie $H = W_k + W_p$.

In der klassischen Statistik werden die Teilchen als unterscheidbar angesehen, wie das für makroskopische Objekte offensichtlich ist. So kann man drei Billardkugeln durch ihre Farben weiß, gelb oder rot auseinanderhalten. Beim Vertauschen zweier Kugeln ergibt sich ein anderer Zustand des Systems. Dabei hat die Farbe keinen Einfluß auf die Stoßeigenschaften, d. h., sie geht nicht in die Bewegungsgleichung der Kugeln ein. Diese Unterscheidbarkeit muß in der Quantenstatistik aufgegeben werden (vgl. 4.10.). Orte und Impulse aller Teilchen zusammen sind insgesamt $6N$ Zahlenangaben. Denkt man sich einen $6N$-dimensionalen Raum, der von den Variablen $p_{1x}, \ldots, p_{Nz}, x_1, \ldots, z_N$ aufgespannt wird, dann wird ein Zustand des Systems durch einen Punkt in diesem Raum beschrieben. Diesen $6N$-dimensionalen Raum nennt man den *Phasen-* oder auch *Γ-Raum*. Die Summe über alle Zustände ν bedeutet in der klassischen Statistik für eine kanonische Gesamtheit eine Integration über den Phasenraum. Die Zustandssumme lautet also

$$Z = (1/N!h^{3N}) \int \mathrm{d}\boldsymbol{p}_1 \ldots \mathrm{d}\boldsymbol{p}_N \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \times \exp[-\beta H(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N; \boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)].$$

Der Vorfaktor $1/(N!h^{6N})$ geht in die Ergebnisse nicht ein. Er wurde so gewählt, daß die Quantenstatistik im klassischen Grenzfall auf die gleiche Formel führt.
Die Impulsintegrationen in Z können ausgeführt werden, da $\exp(-\beta W_k)$ mit $W_k = \sum_i p_i^2/2m$ in gleichartige Faktoren zerfällt und über jede Variable unabhängig von den anderen integriert werden kann. Mit der Abkürzung $z = V \int (\mathrm{d}^3 p/h^3) \exp(-\beta p^2/2m)$ schreibt sich die Zustandssumme

$Z = (1/N!) z^N Q_N, \quad z = V(2\pi mkT/h^2)^{3/2}.$

Die Komplikationen liegen allein bei der Auswertung des *Konfigurationsintegrals*

$Q_N = V^{-N} \int \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N \exp(-\beta W_p),$

das die Wechselwirkung zwischen den Teilchen enthält. Der *Mittelwert* $\bar{A} = \sum_\nu A_\nu w_\nu$ einer physikalischen Größe $A(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N; \boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ ist in der klassischen Statistik aus

$$\bar{A} = \frac{1}{Z} \frac{1}{N!h^{6N}} \int \mathrm{d}\boldsymbol{p}_1 \ldots \mathrm{d}\boldsymbol{p}_N \mathrm{d}\boldsymbol{r}_1 \ldots \mathrm{d}\boldsymbol{r}_N A \exp(-\beta H)$$

zu berechnen.

4.9.1. Gleichverteilungssatz

Wird nach der *mittleren kinetischen Energie* $\bar{\varepsilon}_{\mathrm{kin}}$ eines Teilchens gefragt, dann ist der Mittelwert der Größe $A = \boldsymbol{p}_1^2/2m$ zu bilden. Die Beiträge aller Integrationen in $\bar{A}$, außer der über $\boldsymbol{p}_1$, kürzen sich durch den Faktor $1/Z$ heraus, es verbleibt $\bar{\varepsilon}_{\mathrm{kin}} = \frac{1}{z} \int \frac{\mathrm{d}p_1}{h^3} \frac{p_1^2}{2m} \exp\left(-\beta \frac{p_1^2}{2m}\right) = \frac{3}{2} kT$. Das Ergebnis ist unabhängig davon, welches Teilchen (hier Teilchen 1) herausgegriffen wurde. Der Mittelwert von $p_{1x}^2/2m$ führt auf ein Drittel dieses Wertes. Man erhält als wichtiges Ergebnis der klassischen Statistik den Gleichverteilungssatz: *Die mittlere kinetische Energie je Freiheitsgrad beträgt im thermodynamischen Gleichgewicht $kT/2$.*

4.9.2. Ideales Gas

Beim idealen Gas werden wechselwirkungsfreie Teilchen betrachtet. Ohne äußeres Kraftfeld gilt damit $W_p = 0$. Daraus folgen $Q_N = 1$ und $Z = z^N/N! = (\mathrm{e}z/N)^N$, wobei die Stirlingsche Formel $N! \approx (N/\mathrm{e})^N$ verwendet wurde. Mit $F = -kT \ln Z$ gewinnt man aus Z

$$F(T,V) = -NkT \left\{ \ln \left[\frac{V}{N} \left(\frac{2\pi mkT}{h^2} \right)^{3/2} \right] + 1 \right\},$$

das thermodynamische Potential freie Energie. Für die Zustandsgleichung $p(V,T) = -(\partial F/\partial V) = NkT/V$ und die innere Energie $U(T) = 3NkT/2 = C_V T$ (für $N = L$) folgen daraus die bekannten Ausdrücke. Die Eigenschaften des idealen Gases sind mittels der klassischen Statistik einfach zu berechnen. Für die Molwärme des einatomigen Gases ergibt sich $C_V = (3/2)Lk = 3R/2$.

4.9.3. Reales Gas

Die Wechselwirkung zwischen zwei Teilchen eines Gases werde durch das Potential $w(r)$ (Abb. 4.9.3–1) beschrieben. Für kleine Abstände beobachtet man eine starke Abstoßung, für mittlere Abstände eine Anziehung. Außerhalb der Reichweite R bewegen sich die Teilchen unabhängig voneinander. Beim Modell des *Hartkugelgases* gibt es nur einen abstoßenden, undurchdringlichen Kern, die Teilchen können sich nur bis zum Abstand σ nähern. Die potentielle Energie des Systems ist ohne äußere Felder $W_p = \sum_{i<j} w(|\boldsymbol{r}_i - \boldsymbol{r}_j|)$. Für ein stark verdünntes Gas, d. h. unter der Annahme, daß nicht drei Teilchen gleichzeitig miteinander wechselwirken, kann man das Konfigurationsintegral Q_N auswerten. Schreibt man es in der Form $Q_N = \mathrm{e}^{N\alpha}$, dann erhält man $\alpha = (N/V) \int \mathrm{d}^3\boldsymbol{r}\, [\mathrm{e}^{-w(r)/kT} - 1]$. Die Zustandsgleichung, die aus dieser Zustandssumme folgt, ist vom Typ der van-der-Waals-Gleichung. Die konkrete Form hängt von dem Verlauf von $w(r)$ ab.
Als *Virialentwicklung* bezeichnet man die Entwicklung nach Potenzen der Dichte $\varrho = N/V$. Man schreibt

$$pV/(NkT) = 1 + \sum_{n=1}^{\infty} c_n \varrho^n, \quad c_1 = -1/2 \int \mathrm{d}^3\boldsymbol{r}\, [\mathrm{e}^{-w(r)/kT} - 1]$$

für die Zustandsgleichung. Die Ausgangsnäherung ist die Zustandsgleichung des idealen Gases. In der oben diskutierten Näherung wird der Virialkoeffizient c_1 berücksichtigt. Für das Hartkugelgas wurden auch höhere Virialkoeffizienten berechnet, $c_1 = (4/6)\sigma^3$, $c_2 = (5/8)c_1^2$, $c_3 = 0{,}2869 c_1^3$, $c_4 = 0{,}1103 c_1^4$.

4.9.4. Verteilungsfunktion

Die N-Teilchen eines Systems bewegen sich mit verschiedenen Geschwindigkeiten. Die Wahrscheinlichkeit, ein Teilchen in einem herausgegriffenen Volumenelement des Geschwindigkeitsraumes $\mathrm{d}^3\boldsymbol{v}$ bzw. des Impulsraumes $\mathrm{d}^3\boldsymbol{p}$ anzutreffen, ist im thermodynamischen Gleichgewicht durch die Maxwellsche Geschwindigkeitsverteilung gegeben. Multipliziert mit N, gibt sie die mittlere Anzahl der Teilchen mit einer Geschwindigkeit in $\mathrm{d}^3\boldsymbol{v}$ bzw. $\mathrm{d}^3\boldsymbol{p}$ an. Als Teilchendichte $\varrho(\boldsymbol{r})$ bezeichnet man die mittlere Anzahl von Teilchen in einem Volumenelement des Ortsraumes $\mathrm{d}^3\boldsymbol{r}$.
Zur gleichzeitigen Charakterisierung von Orts- und Geschwindigkeitsverteilung führt man die Verteilungsfunktion $f(\boldsymbol{p}, \boldsymbol{r}, t)$ ein. $f(\boldsymbol{p}, \boldsymbol{r}, t)\, \mathrm{d}^3\boldsymbol{p}\, \mathrm{d}^3\boldsymbol{r}$ soll die Anzahl der Teilchen

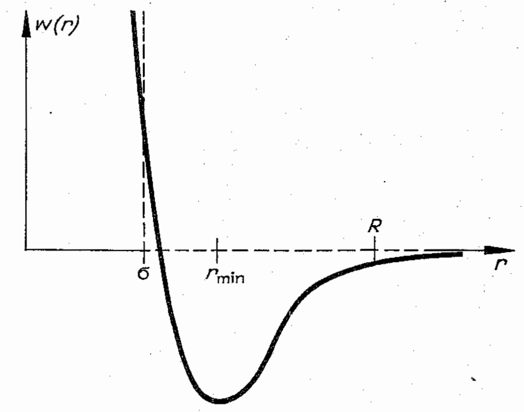


Abb. 4.9.3–1 Qualitativer Verlauf des Potentials der Wechselwirkung zwischen zwei Atomen (– – – Hartkugelgas)

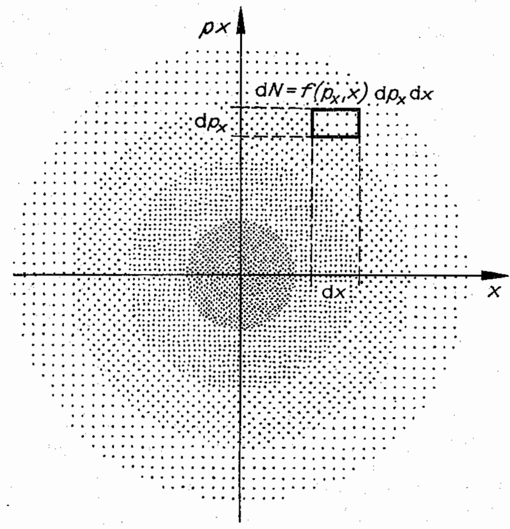


Abb. 4.9.4–1 Phasenraumelement $dp_x dx$ mit dN Teilchen

angeben, die sich zur Zeit t in der Umgebung des Punktes $\boldsymbol{r}$ im Volumenelement d^3r befinden und einen Impuls in einem Volumenelement d^3p um den Wert $\boldsymbol{p}$ herum aufweisen. $d^3p d^3r$ ist ein Volumenelement im sechsdimensionalen Phasenraum, dem sogenannten *μ-Raum.* Für den Fall eines Freiheitsgrades, in dem der Phasenraum zweidimensional ist, werden diese Begriffe in Abb. 4.9.4–1 veranschaulicht. Die Verteilungsfunktion soll entsprechend

$$N = \int d^3p d^3r f(\boldsymbol{p}, \boldsymbol{r}, t), \quad \varrho(\boldsymbol{r}, t) = \int d^3p f(\boldsymbol{p}, \boldsymbol{r}, t)$$

normiert sein. Die Integration über alle Teilchenimpulse liefert die Teilchendichte. Im homogenen System und im stationären oder Gleichgewichtsfall hängt $f(\boldsymbol{p}, \boldsymbol{r}, t)$ nicht von $\boldsymbol{r}$ und t ab. Im thermodynamischen Gleichgewicht ist f durch die Maxwellsche Geschwindigkeitsverteilung (vgl. 2.6.12.2.) gegeben,

$$f(\boldsymbol{p}) = \frac{N}{V}\left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{p^2}{2mkT}\right).$$

Betrachtet man ein System im Schwerefeld, dann ist der Faktor N/V durch die barometrische Höhenformel zu ersetzen.

4.9.5. Paarverteilung

Bei wechselwirkenden Teilchen ist die Bewegung zweier Teilchen korreliert. Die Teilchen weichen sich bei Abstoßung aus, bei Anziehung versuchen sie, sich zusammenzulagern.

Die *Paarverteilung* beinhaltet die Wahrscheinlichkeit $P(\boldsymbol{r}, \boldsymbol{r}')\, d^3r d^3r'$, in einem System zwei Teilchen an den Orten $\boldsymbol{r}$ bzw. $\boldsymbol{r}'$ anzutreffen. Sie ist für eine kanonische Gesamtheit definiert über

$$P^{(N)}(\boldsymbol{r}_1, \boldsymbol{r}_2) = \frac{N(N-1)}{2} \frac{1}{Q^N V^N} \int d^3r_3 d^3r_4 \ldots d^3r_N \exp(-\beta W_p).$$

Die Integration von $P^{(N)}$ über $d^3r_1 d^3r_2$ liefert die Anzahl der Paare $N(N-1)/2$. Bei einer großkanonischen Gesamtheit muß $P^{(N)}$ über verschiedene Teilchenzahlen N gemittelt werden. Die entstehende Paarverteilung wird mit $P(\boldsymbol{r}, \boldsymbol{r}')$ bezeichnet.

In einem Festkörper gibt es langreichweitige Korrelationen. Die Paarverteilung der Atome ist nur von Null verschieden, wenn $\boldsymbol{r} - \boldsymbol{r}'$ gleich einem Gittervektor ist. In einem Gas oder einer Flüssigkeit liegt eine kontinuierliche Verteilung vor. In einem idealen Gas ist P ortsunabhängig, es treten keine Korrelationen auf.

In einem homogenen System hängt die Paarverteilung nur vom Abstand $r = |\boldsymbol{r} - \boldsymbol{r}'|$ ab. Die Teilchendichte $\varrho = \bar{N}_\nu/V = N/V$ ist konstant. Man führt über $P(\boldsymbol{r}, \boldsymbol{r}') = \varrho^2 g(r)/2$ die *Paarkorrelationsfunktion* $g(r)$ ein. Sie ist eine wichtige Eigenschaft eines Systems bei statistischen Betrachtungen, weil aus ihr z. B. *Zustandsgleichung*

$$\frac{pV}{NkT} = 1 - \frac{2\pi}{3kT} \frac{N}{V} \int_0^\infty g(r)\, r^3 \frac{dw}{dr} dr,$$

innere Energie

$$\frac{U}{N} = \frac{3}{2} kT + 2\pi \frac{N}{V} \int_0^\infty g(r)\, r^2 w(r)\, dr \text{ und}$$

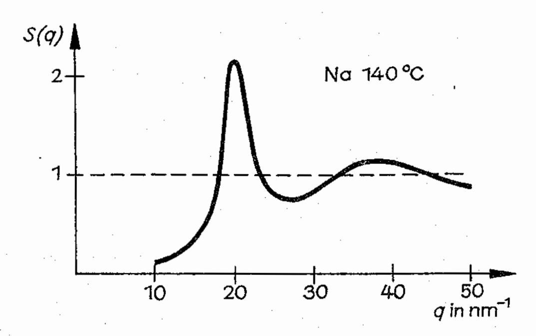
isotherme Kompressibilität $\varkappa_T = -(1/V)\,(\partial V/\partial p)_T$

$$\frac{NkT}{V} \varkappa_T = 1 + \frac{N}{V} \int [g(r) - 1]\, 4\pi r^2 dr$$

berechnet werden können. Für das ideale Gas gelten $w(r) = 0$ und $g(r) = 1$.

Der *Strukturfaktor* hängt über eine Fourier-Transformation mit der Paarkorrelationsfunktion zusammen,

$$S(\boldsymbol{q}) = 1 + \frac{N}{V} \int [g(r) - 1]\, e^{-i\boldsymbol{q}\cdot\boldsymbol{r}} d^3r.$$



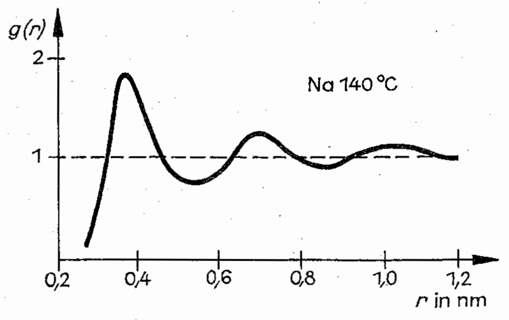


Abb. 4.9.5–1 Strukturfaktor $S(q)$ und Paarverteilung $g(r)$ von flüssigem Natrium

$S(q)$ gewinnt man experimentell mittels Beugung von Röntgen- oder Neutronenstrahlen. Man sieht, daß $S(q \to 0)$ unmittelbar die isotherme Kompressibilität liefert, $S(0) = (N/V)\, kT\varkappa_T$. Abb. 4.9.5–1 zeigt *Paarkorrelationsfunktion und Strukturfaktor einer Flüssigkeit.* Aufgrund der Abstoßung zwischen den Teilchen trifft man keine Paare mit Abständen an, die kleiner als der Durchmesser des abstoßenden Kerns im Potential sind. Die folgenden Maxima deuten auf eine *Nahordnung* hin. Für große Abstände geht $g \to 1$, es gibt in der Flüssigkeit keine *Fernordnung.* Die Paarkorrelationsfunktion hängt von der Temperatur ab. Mit sinkender Temperatur werden die Maxima stärker ausgeprägt und bis zu größeren Abständen beobachtbar. Beim Erstarren zum geordneten Festkörper gehen die Maxima in scharfe Spitzen über, deren Position den Gitterabständen im Kristall entsprechen.

4.9.6. Boltzmann-Gleichung

Als Boltzmann-Gleichung bezeichnet man die Bilanzgleichung für die Verteilungsfunktion $f(\boldsymbol{p}, \boldsymbol{r}, t)$. Betrachtet wird ein Volumenelement im Ortsraum $d^3\boldsymbol{r}$. Die mittlere freie Weglänge (vgl. 2.1.12.) soll klein gegen die Ausdehnung dieses Volumenelementes sein. Untersucht werden zur Zeit t diejenigen Teilchen in diesem Volumenelement, deren Impuls innerhalb von $d^3\boldsymbol{p}$ liegt. Der Definition der Verteilungsfunktion entsprechend ist die Anzahl dieser Teilchen $dN = f(\boldsymbol{p}, \boldsymbol{r}, t) d^3\boldsymbol{p} d^3\boldsymbol{r}$. dN kann sich durch drei Prozesse ändern: 1. Wenn ein Dichteunterschied zu benachbarten Volumenelementen besteht, dann ändert sich die Teilchenzahl durch *Diffusion.* 2. Durch äußere *Kraftfelder* werden die Teilchen beschleunigt oder abgebremst, ihr Impuls wandert aus $d^3\boldsymbol{p}$ heraus. Man kann dies auch als Diffusion im Impulsraum bezeichnen. 3. *Stöße* zwischen den Teilchen führen zu einer Impulsänderung. Die Teilchen werden an beliebige Stellen des Impulsraumes gestreut. Die Bilanzgleichung hat also die Form

$$\partial f/\partial t + (\partial f/\partial t)_{\text{Diff.}} + (\partial f/\partial t)_{\text{Kraft}} + (\partial f/\partial t)_{\text{Stoß}} = 0.$$

Bei allen drei Prozessen treten Ereignisse auf, bei denen sich die Teilchenzahl im betrachteten Phasenraumelement $d^3\boldsymbol{p} d^3\boldsymbol{r}$ vermindert wie auch solche, bei denen sie sich erhöht. Es wird angenommen, daß keine Prozesse auftreten, bei denen Teilchen erzeugt oder vernichtet werden. Sonst treten Zusatzterme auf. Bei der Neutronentransportgleichung, die den Neutronenfluß in Kernreaktoren bestimmt, muß dies z. B. beachtet werden, weil der Einfang von Neutronen durch Uraniumatome und das Freiwerden von Neutronen bei der Kernspaltung wesentliche Anteile in der Bilanz darstellen.

Ein Stoßprozeß zwischen zwei Teilchen (Abb. 4.9.6–1) sei durch eine Übergangswahrscheinlichkeit $Q(\boldsymbol{p}, \boldsymbol{p}_2; \boldsymbol{p}_1', \boldsymbol{p}_2')$ charakterisierbar. Q wird durch die Form der Kräfte zwischen den Teilchen bestimmt und ist bis auf Faktoren gleich dem *Wirkungsquerschnitt* (vgl. 2.1.10.4.). Unter der Voraussetzung, daß die Reichweite der zwischenatomaren Kräfte klein gegen die mittlere freie Weglänge ist, leitete BOLTZMANN die folgende *Bilanzgleichung für die Verteilungsfunktion* $f(\boldsymbol{p}, \boldsymbol{r}, t)$ ab:

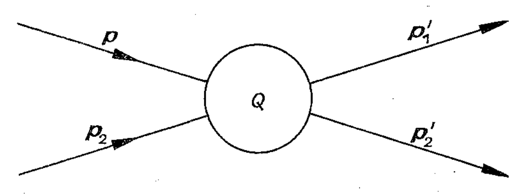


Abb. 4.9.6–1 Impulsänderung beim Stoß

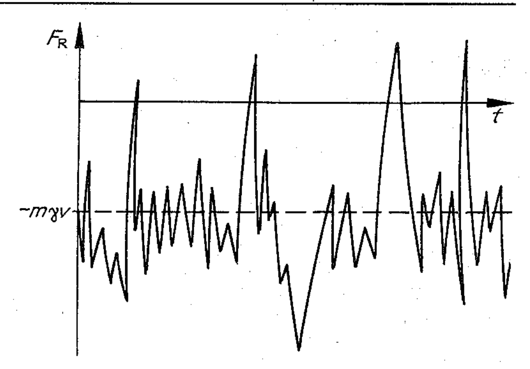


Abb. 4.9.7–1 Stochastische Kraft

$$\partial f/\partial t = -\boldsymbol{p}\,\partial f/\partial \boldsymbol{r} - \boldsymbol{F}\partial f/\partial \boldsymbol{p} - \int d^3\boldsymbol{p}_2 d^3\boldsymbol{p}_1' d^3\boldsymbol{p}_2' \,[f(\boldsymbol{p}) f(\boldsymbol{p}_2) - f(\boldsymbol{p}_1') f(\boldsymbol{p}_2')]\, Q(\boldsymbol{p}, \boldsymbol{p}_2; \boldsymbol{p}_1', \boldsymbol{p}_2').$$

In dem Stoßterm beschreibt der erste Summand die Teilchen, die aus dem betrachteten Impulsbereich herausgestreut werden, der zweite Summand die Teilchen, die hineingestreut werden. Im thermodynamischen Gleichgewicht sind diese beiden Anteile gleich groß. Man erkennt dies, wenn man die Gleichgewichtsverteilung $f(\boldsymbol{p}) \sim \exp(-\boldsymbol{p}^2/2mkT)$ einsetzt und beachtet, daß bei elastischen Stößen $\boldsymbol{p}^2 + \boldsymbol{p}_2^2 = \boldsymbol{p}_1'^2 + \boldsymbol{p}_2'^2$ gilt.

Das H-Theorem leitete BOLTZMANN unter Verwendung der obigen Bilanzgleichung ab. Er betrachtete den Ausdruck $H = \int d^3\boldsymbol{p} f(\boldsymbol{p}) \ln f(\boldsymbol{p})$ und konnte zeigen: Es gilt $(\partial H/\partial t) \leqq 0$, wenn $f(\boldsymbol{p})$ der Boltzmann-Gleichung genügt. Das Gleichheitszeichen tritt im thermodynamischen Gleichgewicht auf. $-H$ wächst also im Zeitablauf ständig an und erreicht im thermodynamischen Gleichgewicht seinen Maximalwert. Dieses Verhalten zeigt auch die Entropie. Dadurch wird der in der Statistik verwendete Ausdruck für die Entropie (vgl. 4.8.1.) nahegelegt.

4.9.7. Fluktuationen

Im thermodynamischen Gleichgewicht treten Schwankungen physikalischer Größen um die Mittelwerte auf. Diese sind zwar klein (vgl. 4.8.4.), geben aber Anlaß zu bestimmten meßbaren Effekten, z. B. Diffusion. Betrachtet wird ein Teilchen in einem thermodynamischen System bzw. eine Schar solcher Teilchen. Auf dieses Teilchen möge eine mit der Kreisfrequenz periodische äußere Kraft $F(\omega)$ wirken. Weiterhin wirkt durch die Stöße mit den Teilchen des Systems auf das herausgegriffene Teilchen eine unregelmäßige (stochastische) Kraft F_R. Diskutiert wird nur der Fall, daß diese Kraft im Mittel proportional zur Geschwindigkeit und bremsend ist (Abb. 4.9.7–1), $F_R = -m\gamma v + F_S(t)$. Definitionsgemäß sollen Zeitmittelwert $\overline{F_S^t} = 0$ über Zeiten, die groß gegen die mittlere Stoßzeit $t \gg \tau_s$ sind, und Scharmittelwert $\langle F_S(t) \rangle = 0$ über die *stochastische Kraft* F_S verschwinden.

Als Bewegungsgleichung für das betrachtete Teilchen kann die *Langevinsche Gleichung* $\dot{v} = F/m - \gamma v + F_S/m$ angesehen werden. Als *Beweglichkeit* des Teilchens bezeichnet man den Proportionalitätsfaktor zwischen äußerer Kraft und mittlerer Geschwindigkeit $\bar{v} = B(\omega)F(\omega)$. Für die betrachtete, geschwindigkeitsproportionale Reibungskraft folgt im statischen Fall $\omega = 0$ aus der Langevinschen Gleichung $B = 1/m\gamma$. Unter Verwendung des Gleichverteilungssatzes $m\langle\dot{x}^2\rangle/2 = kT/2$ erhält man aus der Langevinschen Gleichung für die *Diffusion* $\langle x^2\rangle = 2BkTt$, also den Zusammenhang $D = BkT$ zwischen Diffusionskoeffizient D und Beweglichkeit B.

Korrelationsfunktionen. Durch die stochastischen Stöße ist der Zustand eines Teilchens zu einer Zeit $t' \gg t$ völlig unabhängig davon, in welchem Zustand sich das Teilchen zur Zeit t befunden hatte. Über welchen Zeitraum $t' - t$ noch ein Zusammenhang besteht, wie gut das „Gedächtnis" des Systems ist, wird durch Zeit-Korrelationsfunktionen beschrieben. Für die Geschwindigkeit ergibt sich aus der Langevinschen Gleichung $\langle v(t)\, v(t')\rangle = (kT/m)\,\mathrm{e}^{-\gamma|t'-t|}$. Für $t' = t$ entspricht der Wert dem Gleichverteilungssatz. Die Zeit $\tau = 1/\gamma$ charakterisiert das Gedächtnis des Systems, für Zeiten $|t' - t| \gg \tau$ gibt es keine Korrelationen mehr. In dem Spezialfall, daß $\tau = 0$ gilt, spricht man von *Markowschen Prozessen.*

Dissipations-Fluktuations-Theorem. Die Beweglichkeit kann aus der Fourier-Transformierten der Geschwindigkeitskorrelationsfunktion bestimmt werden,

$$B(\omega) = (1/kT) \int_0^\infty \mathrm{d}t \langle v(t_0)\, v(t_0 + t)\rangle\, \mathrm{e}^{-i\omega t}.$$

Das dissipative Verhalten eines Systems (z. B. die auftretende Reibungskraft, hier beschrieben durch die Beweglichkeit) beim Einfluß äußerer Felder, also die Reaktion des Systems auf eine äußere Störung, wird durch die Fluktuationseigenschaften des ungestörten Systems um den thermodynamischen Gleichgewichtszustand (hier durch die Geschwindigkeitskorrelationsfunktion) bestimmt. Ein einfaches Beispiel dafür ist der Zusammenhang zwischen Beweglichkeit und Diffusionskoeffizient. Im Fall einer geschwindigkeitsproportionalen mittleren Reibungskraft führt obiger Zusammenhang auf $B(\omega) = 1/[m(\gamma + i\omega)]$.

4.10. Quantenstatistik

Die Quantenmechanik liefert als wesentlichen Unterschied zur klassischen Mechanik *diskrete Werte* für die Erhaltungsgrößen, z. B. für die Energie. Beschränkt man sich auf die Betrachtung *wechselwirkungsfreier Teilchen* – gegebenenfalls im Sinne der Hartree-Fock-Näherung – dann gibt es Einteilchenniveaus ε_i, z. B. die Energieniveaus eines Atoms. Ein Zustand eines Systems von N Teilchen ist dann dadurch gekennzeichnet, wie die N Teilchen auf die Niveaus ε_i verteilt sind. Mit $n_i^{(\nu)}$ wird die Anzahl der Teilchen im Niveau i für den Zustand ν bezeichnet. Die Gesamtzahl der Teilchen $\sum_i n_i^{(\nu)} = N$ hat immer den Wert N. Die in der statistischen Mechanik für eine *kanonische Gesamtheit* auftretende Summe über alle Zustände bedeutet im Falle der Quantenstatistik also die Summe über alle Möglichkeiten, N Teilchen auf die diskreten Niveaus ε_i zu verteilen.

Die *Ununterscheidbarkeit der Elementarteilchen* bedingt die Unterschiede der quantenmechanischen Zählweise zur klassischen Zählweise. Nach der klassischen Zählweise sind zu unterscheiden: 1. das Teilchen 1 befindet sich im Niveau ε_i, das Teilchen 2 befindet sich im Niveau ε_j und 2. das Teilchen 1 befindet sich im Niveau ε_j, das Teilchen 2 befindet sich im Niveau ε_i. Beide Zustände müssen in der Summe $\sum_\nu$ erfaßt werden. Wegen der Ununterscheidbarkeit kann man in der Quantenstatistik lediglich feststellen, daß sich in den Niveaus ε_i und ε_j jeweils ein Teilchen befindet, die beiden oben genannten Fälle können nicht unterschieden werden. Der quantenmechanische Zustand ist durch die Angabe der $n_i^{(\nu)}$ bestimmt. Eine Angabe, welches Teilchen sich in welchem Niveau befindet, ist nicht möglich.

Die *Fermi-Dirac-* bzw. *Bose-Einstein-Statistik* drückt weitere Besonderheiten gegenüber der klassischen Statistik aus. Es sind zwei Arten von Elementarteilchen bekannt, Fermionen und Bosonen (vgl. 4.7.). Sie unterscheiden sich in den möglichen Werten der *Besetzungszahl* n_i eines Einteilchenzustandes i: $n_i = 0,1$ für Fermionen, $n_i = 0,1,2,\ldots$ für Bosonen. Bei Fermionen kann ein Zustand also nur unbesetzt oder besetzt, aber nicht mehrfach besetzt sein. Diese Aussage wird auch als *Pauli-Prinzip* bezeichnet. Abb. 4.10.0–1 illustriert die Anzahl der Zustände eines Systems zweier Teilchen, die auf drei Niveaus verteilt werden können. Die Zahl der Summanden z. B. in der Zustandssumme ist also in den drei Fällen voneinander verschieden, was auch zu wesentlichen Unterschieden in den physikalischen Eigenschaften führt.

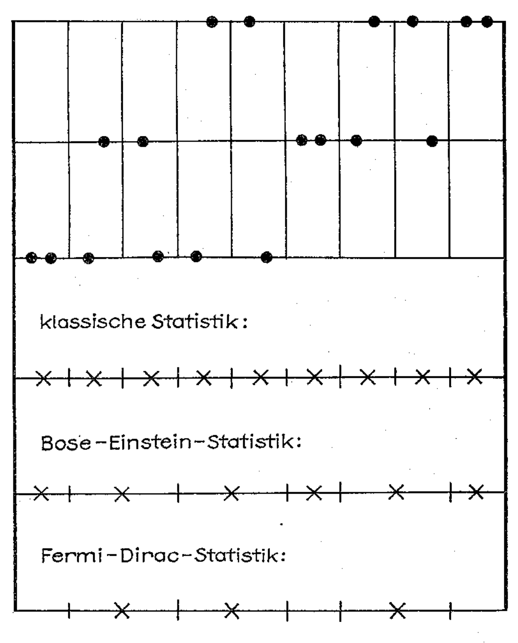


Abb. 4.10.0–1 Unterschiedliche Anzahl von Zuständen bei klassischen Teilchen, Bosonen bzw. Fermionen

Quasikontinuierliche Spektren treten in der Quantenphysik neben den diskreten Spektren auf. Ein freies Teilchen, das sich in einem Volumen V aufhält, ist durch seinen Impuls $\boldsymbol{p}$ gekennzeichnet. Die Bohr-Sommerfeldsche Quantenbedingung besagt, daß sich die Phasenfläche je Freiheitsgrad von Niveau zu Niveau um h ändert. Bei drei

Freiheitsgraden nimmt ein Zustand das Phasenraumvolumen $V\Delta^3 p = h^3$ ein. Im Impulsraum liegt also je Element $\Delta^3 p = h^3/V$ ein Zustand. Da V eine makroskopische Größe ist, sind das sehr kleine Impulsvolumenelemente. Anstatt über diese Elemente zu summieren, wird man eine Integration durchführen, $\sum_p$ wird durch $\int \mathrm{d}^3 p V/h^3$ ersetzt. Bei N Teilchen hat ein Phasenraumelement das Volumen h^{3N}. Integriert man alle Impulsvariablen $p_1, \ldots, p_N$ über den gesamten Impulsraum, dann hat man Zustände, die durch Vertauschen von Teilchen entstehen, getrennt gezählt. Das widerspricht der Unterscheidbarkeit, so daß mit einem Faktor $1/N!$ multipliziert werden muß.

4.10.1. Paramagnetismus

Ein magnetischer Dipol μ hat in einem Magnetfeld die *Wechselwirkungsenergie* $E = -\mu B$. Energetisch am günstigsten ist die Parallelstellung zum Feld. Steht der Dipol in Kontakt mit einem Wärmebad, dann wird das Auftreten einer bestimmten Orientierung durch die Boltzmann-Wahrscheinlichkeit $w \sim \mathrm{e}^{-E/kT}$ bestimmt. Bei klassischer Betrachtung kann der Dipol beliebig orientiert sein. Bei quantenmechanischer Betrachtung ist das magnetische Moment mit einem Drehimpuls verbunden, der nur *diskrete Einstellmöglichkeiten* aufweist. Wird die z-Achse in Richtung des Magnetfeldes $B = Be_z$ gelegt, dann tritt in der Wechselwirkungsenergie nur die z-Komponente des magnetischen Momentes auf. Betrachtet man den Fall, daß das Teilchen den Spin 1/2 aufweist, dann hat der Spin und damit das magnetische Moment zwei Einstellmöglichkeiten. Die zugehörigen Erwartungswerte der z-Komponente des magnetischen Momentes sollen mit $\pm\mu_z$ bezeichnet werden. Der Dipol hat im Magnetfeld zwei Zustände mit den Energien $\pm\mu_z B$. Die Zustandssumme besteht also nur aus zwei Summanden $z = \exp(\mu_z B/kT) + \exp(-\mu_z B/kT)$. Bei einem System von N unabhängigen Dipolen ist die Zustandssumme $Z = z^N$. Über $F = -kT \ln Z$ bzw. $M = -(\partial F/\partial B)$ gewinnt man die freie Energie bzw. das *magnetische Moment* des Systems
$M = N\mu_z \tanh[\mu_z B/(kT)]$.
Für $\mu_z B \gg kT$ geht der tanh gegen Eins. Bei hoher Feldstärke oder tiefen Temperaturen befinden sich alle Dipole in dem energetisch günstigsten Zustand $M = N\mu_z$. Für schwache Felder $\mu_z B \ll kT$ wird das magnetische Moment

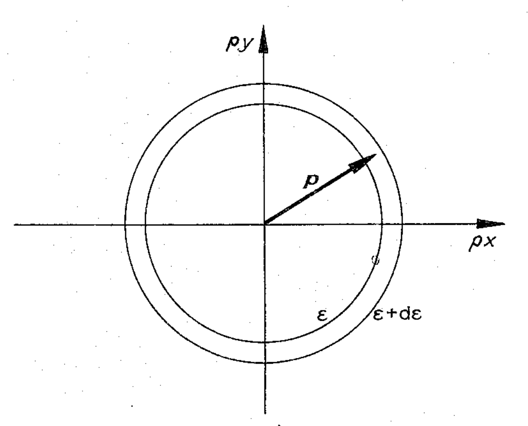


Abb. 4.10.2–1 Zustände innerhalb von dε

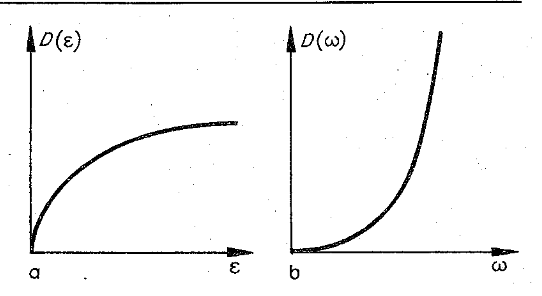


Abb. 4.10.2–2 Zustandsdichte von (a) freien Elektronen, (b) Photonen oder Phononen

im thermodynamischen Gleichgewicht $M = N(\mu_z^2/kT)B$ proportional zum angelegten Feld.
Die *klassische Betrachtung* liefert mit $\mu = |\mu|$ für das magnetische Moment $M = N\mu[\coth(\mu B/kT) - kT/\mu B]$. Die Werte in den Grenzfällen sind $N\mu$ bzw. $N(\mu^2/3kT)B$. Das Ergebnis für schwache Felder stimmt mit dem quantenmechanisch gewonnenen überein, wenn man $\mu_z^2 = \mu^2/3$ setzt.
Die *paramagnetische Suszeptibilität* $\chi_m \sim 1/T$ ist temperaturabhängig und fällt mit wachsender Temperatur ab. Berücksichtigt man, daß das Magnetfeld, welches ein Dipol spürt, auch einen Beitrag von den umgebenden Dipolen enthält, dann gewinnt man das sogenannte Curie-Weiß-Gesetz $\chi_m \sim 1/(T - T_c)$. Unterhalb der Temperatur T_c tritt ohne äußeres Magnetfeld eine Ordnung der magnetischen Dipole ein, wie sie z. B. als Ferromagnetismus bekannt ist.

4.10.2. Zustandsdichte

Der Zustand eines *freien Teilchens* ist durch seinen Impuls p und seine Spineinstellung m_s gekennzeichnet. Die möglichen Impulswerte bilden ein quasikontinuierliches Spektrum, die Summe über alle möglichen Zustände wird daher als ein Integral über den Impulsraum geschrieben. Entsprechend der Quantelung des Phasenraumes ist die Zahl der möglichen Impulswerte in einem Volumenelement $\mathrm{d}^3 p$ durch $\mathrm{d}N = V\mathrm{d}^3 p/h^3$ gegeben. Der Spin hat $g = 2s + 1$ Einstellmöglichkeiten, für Elektronen ist $g = 2$.
Die Zustände mit einer Energie zwischen ε und $\varepsilon + \mathrm{d}\varepsilon$ gehören zu Impulsen, die im Impulsraum innerhalb einer Kugelschale liegen (Abb. 4.10.2–1). Ihre Anzahl wird unter Einführung der *Zustandsdichte* $D(\varepsilon)$ in der Form $D(\varepsilon)\mathrm{d}\varepsilon$ geschrieben.
Ist die Energie der Teilchen durch $\varepsilon(p) = p^2/2m$ gegeben, dann führt die Quantelung des Phasenraumes auf
$D(\varepsilon) = g2\pi V(2m/h^2)^{3/2}\sqrt{\varepsilon}$.
Mit $g = 2$ ist das die Zustandsdichte *freier Elektronen* (Abb. 4.10.2–2a).
Für *Photonen* und *Phononen* gilt $\varepsilon = \hbar\omega = c|p|$ als Zusammenhang zwischen Energie und Impuls. Darin ist c die Lichtgeschwindigkeit bzw. die Schallgeschwindigkeit. Schreibt man die Anzahl der Zustände im Intervall $\mathrm{d}\omega$ in der Form $D(\omega)\mathrm{d}\omega$, dann ergibt sich

$$D(\omega) = [gV/(2\pi^2 c^3)]\,\omega^2$$

(Abb. 4.10.2–2b). Entsprechend den verschiedenen Polarisationsmöglichkeiten ist für Photonen $g = 2$ und für Phononen $g = 3$.

4.10.3. Mittlere Besetzungszahl

Betrachtet wird ein System mit den Einteilchenniveaus ε_i. Ein Zustand ν mit N Teilchen ist dann durch die Besetzungszahlen $n_i^{(\nu)}$ gekennzeichnet. Abb. 4.10.3–1 zeigt ein Beispiel dafür, wie zehn Teilchen auf sieben Niveaus verteilt sein können. Aus den Besetzungszahlen ergeben sich Teilchenzahl N_ν und Energie E_ν des betrachteten Zustandes

$$N_\nu = \sum_i n_i^{(\nu)}, \quad E_\nu = \sum_i n_i^{(\nu)} \varepsilon_i.$$

Im thermodynamischen Gleichgewicht ist der Zustand ν mit der Wahrscheinlichkeit w_ν realisiert. w_ν ist entsprechend der betrachteten Gesamtheit zu wählen. Die *mittlere Teilchenzahl* im Niveau ε_i ergibt sich damit zu $n(\varepsilon_i, T) = \sum_\nu n_i^{(\nu)} w_\nu$. *Die innere Energie* $U = \sum_\nu w_\nu E_\nu$ kann damit auch in der Form

$$U = \sum_i n(\varepsilon_i, T)\, \varepsilon_i = \int d\varepsilon\, D(\varepsilon)\, n(\varepsilon, T)$$

geschrieben werden. Aus der Zustandssumme $Z = \sum_\nu \exp[-\beta(E_\nu - \mu N_\nu)]$ ergibt sich die mittlere Besetzungszahl über $n(\varepsilon_i, T) = (1/\beta)\, (\partial \ln Z/\partial \varepsilon_i)$.

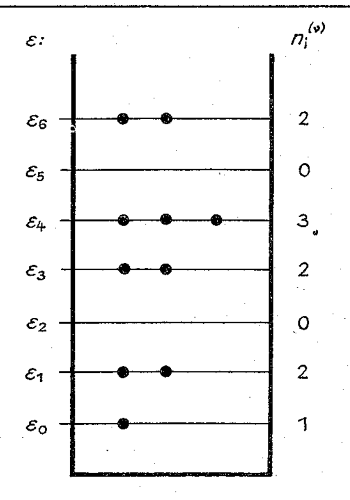


Abb. 4.10.3–1 Besetzungszahlen eines möglichen Zustandes ν

4.10.4. Bose-Einstein-Verteilung

Die *mittlere Besetzung* $n(\varepsilon, T)$ eines Niveaus ε im thermodynamischen Gleichgewicht für *Bosonen* ergibt sich, wenn man als Besetzungszahlen $n_\varepsilon^{(\nu)}$ alle Werte 0, 1, 2, ... zuläßt. Als System betrachte man das einzelne interessierende Niveau ε, das mit einem Reservoir in Kontakt stehen soll und Energie und Teilchen austauschen kann. Die verschiedenen Zustände des Systems sind durch die Besetzungszahl $n_\varepsilon^{(\nu)} = \nu$ klassifizierbar. Da das eine Niveau betrachtet wird, gilt $N_\nu = \nu$ und $E_\nu = \nu\varepsilon$. Damit ergibt sich für die Zustandssumme

$$Z = \sum_{\nu=0}^{\infty} \exp[-\beta(E_\nu - \mu N_\nu)] = 1/\{1 - \exp[-\beta(\varepsilon - \mu)]\}$$

und für die mittlere Besetzungszahl

$$n(\varepsilon, T) = \frac{1}{e^{\beta(\varepsilon-\mu)} - 1},$$

$\beta = 1/kT$, die sogenannte Bose-Einstein-Verteilung (Abb. 4.10.4–1).

Das *chemische Potential* $\mu(T, N)$ ist durch die vorgegebene mittlere Teilchenzahl N bestimmt, $\sum_\varepsilon D(\varepsilon)\, n(\varepsilon, T) = N$. Da $n(\varepsilon, T)$ nicht negativ sein darf – negative Teilchenzahlen sind physikalisch unsinnig –, muß $\mu < \varepsilon_0$ gelten, wenn ε_0 das unterste Einteilchenniveau ist. Für hohe Temperaturen $T \to \infty$ geht $\mu \to -\infty$ derart, daß $e^{-\mu/kT} \gg 1$ gilt. Dann geht die Bose-Einstein-Verteilung in die Boltzmann-Verteilung $n(\varepsilon, T) \sim e^{-\varepsilon/kT}$ über.

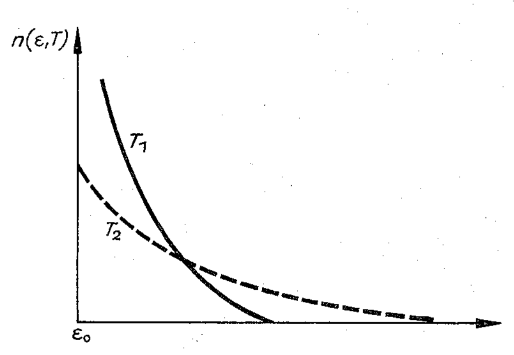


Abb. 4.10.4–1 Qualitativer Verlauf der Bose-Einstein-Verteilung, $T_1 < T_2$

4.10.5. Schwingungsenergie fester Körper

4.10.5.1. Lineare harmonische Oszillatoren haben die äquidistanten Energieniveaus $\varepsilon_n = \hbar\omega(n + 1/2)$. Steht ein solcher Oszillator mit einem Wärmebad in Kontakt, dann ergibt sich aus $Z = \sum_{n=0}^{\infty} \exp(-\beta\varepsilon_n)$ seine *mittlere Energie* zu

$$\varepsilon(\omega, T) = \hbar\omega \left[\frac{1}{2} + \frac{1}{e^{\hbar\omega/kT} - 1}\right] \to \begin{cases} \hbar\omega(1/2 + e^{-\hbar\omega/kT}) \\ kT \end{cases}$$

für $kT \ll \hbar\omega$ bzw. $kT \gg \hbar\omega$.

Diesen Ausdruck – ohne die Nullpunktsenergie $\hbar\omega/2$ – verwendete PLANCK bei der Herleitung seiner Strahlungsformel. Ihre Ähnlichkeit zur Bose-Einstein-Verteilung legt die Schreibweise $\varepsilon(\omega, T) = \hbar\omega\, [1/2 + n_{\text{Phonon}}(T)]$ nahe. Bezeichnet man jede Anregung des Oszillators $\hbar\omega$ als ein Phonon, dann ist $n_{\text{Phonon}}(T)$ die mittlere Zahl der Photonen im thermodynamischen Gleichgewicht, wobei die Phononen Bose-Teilchen sind.

4.10.5.2. Schwingungsenergie eines Festkörpers ergibt sich, wenn man die zu den einzelnen Eigenfrequenzen gehörende mittlere Schwingungsenergie addiert, $U = \int d\omega\, D(\omega)\, \varepsilon(\omega, T)$. Im Debye-Modell (vgl. 6.4.2.) genügen die Schwingungszweige der einfachen Dispersionsbeziehung $\omega = c_s |k|$. Läßt man die Unterschiede in der Ausbreitungsgeschwindigkeit longitudinaler und transversaler Schwingungen außer Betracht, dann ist die Zustandsdichte (vgl. 4.10.2.) $D(\omega) = (3V/2\pi^2 c_s^3)\omega^2 = 9N\omega^2/\omega_D^3$. Debye-Frequenz ω_D und Debye-Temperatur Θ_D hängen über

$\hbar\omega_D = k\Theta_D$ zusammen. Setzt man $D(\omega)$ und $\varepsilon(\omega, T)$ in den Ausdruck für die innere Energie ein, dann erhält man für *hohe Temperaturen* $U = 3NkT$. Das steht in Einklang mit der *Dulong-Petitschen Regel* für die Molwärme $C_V = 3R$ (vgl. 2.6.5.2.). Für $T = 0$ verbleibt die Nullpunktsenergie der $3N$ Eigenschwingungen des Festkörpers $U_0 = 3N(3/4)\hbar\omega_D$. Für *niedrige Temperaturen* $T \ll \Theta_D$ ergibt sich zu U_0 der Zusatzterm $U = U_0 + (3\pi^4/5)Nk\Theta_D(T/\Theta_D)^4$. Die Molwärme $C_V = (\partial U/\partial T) \sim T^3$ ist also der *dritten Potenz der Temperatur* proportional.

4.10.6. Bose-Einstein-Kondensation

Bei tiefen Temperaturen weist ein System von Bose-Teilchen Besonderheiten auf, aus denen man z. B. die *Superfluidität* von ^{4}He, die unterhalb von 2,17 K beobachtet wird, erklären kann. Sie hängen damit zusammen, daß der Nenner in der Bose-Einstein-Verteilung für $\mu = \varepsilon_0$ verschwindet.
Für $T = 0$ befindet sich das betrachtete System von N wechselwirkungsfreien Bosonen im Grundzustand. Der energetisch tiefste Zustand ist dann realisiert, wenn sich alle N Bosonen im tiefsten Niveau ε_0 befinden $n(\varepsilon_0, T = 0) = N$ und alle höher liegenden Niveaus $\varepsilon_1, \varepsilon_2, \ldots$ unbesetzt sind $n(\varepsilon_i, T = 0) = 0$ für $i \neq 0$. Die makroskopische Besetzung des Niveaus ε_0 erfordert $\lim\limits_{T\to 0} \frac{\varepsilon_0 - \mu}{kT} = 1/N$. Das *chemische Potential* rückt außerordentlich dicht an ε_0 heran, der Abstand ist $\sim 1/N$ (Abb. 4.10.6–1).

ε_1

ε_0

μ

Abb. 4.10.6–1 Lage des chemischen Potentials bei $T \to 0$

Die Energieniveaus freier Teilchen in einem makroskopischen Volumen V bilden ein Quasikontinuum. In einem Potentialkasten der Ausdehnung d ist der Abstand $\varepsilon_1 - \varepsilon_0 \approx \hbar^2/2md^2 \sim V^{-2/3}$. Wird ein System konstanter Dichte betrachtet, dann ist $V \sim N$, also $\varepsilon_1 - \varepsilon_0 \sim N^{-2/3}$. Da N eine makroskopische Zahl von der Größenordnung 10^{24} ist, sind diese Energieabstände winzig, gemessen in atomaren Einheiten. Der Abstand $\varepsilon_0 - \mu$ ist aber noch viel kleiner, $(\varepsilon_0 - \mu)/(\varepsilon_1 - \varepsilon_0) \sim N^{-1/3} \approx 10^{-8}$ (vgl. Abb. 4.10.6–1). Diesen Sachverhalt muß man vor Augen haben, wenn man das Phänomen der Bose-Einstein-Kondensation aus der Verteilungsfunktion heraus verstehen will.
In dem Temperaturbereich, in dem $(\varepsilon_0 - \mu)/(\varepsilon_1 - \varepsilon_0) \ll 1$ ist, ist die Besetzung eines Zustandes $i \neq 0$ wegen $(\varepsilon_i - \mu)/(kT) \approx (\varepsilon_i - \varepsilon_0)/(kT)$ durch $n(\varepsilon_i, T) = 1/(e^{\beta(\varepsilon_1 - \varepsilon_0)} - 1)$ gegeben. Dieser Ausdruck ist aber unabhängig von der Gesamtteilchenzahl N; N kommt ja nur über das chemische Potential in die Verteilungsfunktion hinein. Verändert man die Teilchenzahl (Abb. 4.10.6–2), dann wird nur die Besetzung von ε_0 geändert, die Besetzungszahl aller anderen Niveaus bleibt unverändert. In diesem Temperaturbereich bleibt

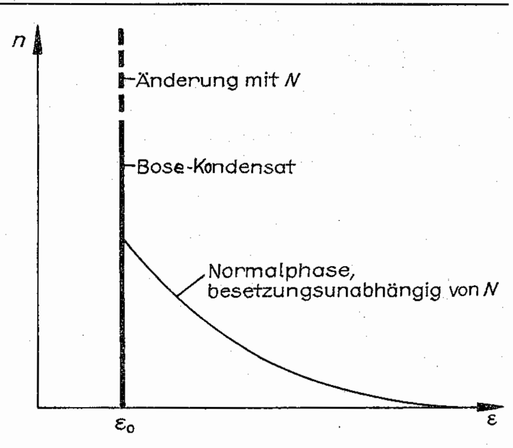


Abb. 4.10.6–2 Bose-Kondensat und Normalphase

also das Niveau ε_0 mit einer makroskopischen Zahl von Bosonen besetzt. Es ist daher berechtigt, von zwei Phasen des Systems zu sprechen, dem Bose-Kondensat – das ist die superfluide Phase – und einer normalen Phase.
Aus dem Ausdruck für die Gesamtteilchenzahl $N = N_0(T) + \int_{\varepsilon_1}^{\infty} d\varepsilon\, D(\varepsilon)\, n(\varepsilon, T) \equiv N_s + N_{\text{normal}}$ berechnet sich $N_{\text{normal}} = 2{,}612\,(mkT/2\pi\hbar^2)^{3/2}V$. Die Temperaturabhängigkeit der Anzahl der Teilchen in der superfluiden Phase ist also von der Form $N_s = N(1 - (T/T_0)^{3/2})$. Diese Betrachtungen sind für $T < T_0 = (2\pi\hbar^2/mk)\,(N/2{,}612V)^{2/3}$ berechtigt. Mit den Werten für ^{4}He ergibt sich $T_0 \approx 3$ K. Das liegt nahe bei dem Umwandlungspunkt 2,17 K zur superfluiden Phase. Die Abweichungen zu dieser Abschätzung liegen darin begründet, daß die ^{4}He-Atome keine wechselwirkungsfreien Teilchen sind.

4.10.7. Fermi-Dirac-Verteilung

Die *mittlere Besetzungszahl* $n(\varepsilon, T)$ eines Niveaus ε im thermodynamischen Gleichgewicht *für Fermionen* kann nur zwischen Null und Eins liegen, da die Besetzungszahl $n_\varepsilon^{(\nu)}$ nur die Werte Null und Eins annimmt; das Niveau ist unbesetzt oder besetzt, es kann nicht mehrfach besetzt sein. Bei einer zu 4.10.4. analogen Betrachtung treten in der Zustands-

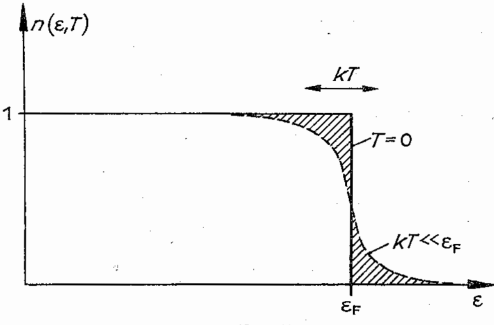


Abb. 4.10.7–1 Fermi-Dirac-Verteilung

summe nur die beiden Summanden zu $\nu = 0,1$ auf $Z = 1 + e^{-\beta(\varepsilon-\mu)}$. Daraus folgt als mittlere Besetzungszahl die sogenannte Fermi-Dirac-Verteilung

$$n(\varepsilon, T) = 1/(e^{\beta(\varepsilon-\mu)} + 1),$$

$\beta = 1/(kT)$. Da der Nenner $\geqq 1$ ist, kann n nicht größer als Eins werden. Für $T \to 0$ hängt der Wert von n wesentlich vom Vorzeichen von $\varepsilon - \mu$ ab. Für $\varepsilon > \mu$ wird $n = 0$, für $\varepsilon < \mu$ erhält man $n = 1$. Das chemische Potential $\mu(T = 0)$ nennt man die *Fermi-Energie* ε_F. Im Grundzustand ($T = 0$) sind also alle unter der Fermi-Energie liegenden Niveaus besetzt, alle darüberliegenden unbesetzt (Abb. 4.10.7–1). Für endliche Temperaturen mit $kT \ll \varepsilon_F$ wird der Sprung der Verteilungsfunktion an der Fermi-Energie verwischt, jedoch nur innerhalb eines Bereiches der Breite kT. Für hohe Temperaturen $kT \gg \varepsilon_F$ geht die Fermi-Dirac-Verteilung – wie die Bose-Einstein-Verteilung – in die Boltzmann-Verteilung $n \sim e^{-\varepsilon/kT}$ über. Für die Leitungselektronen in einem Metall liegt ε_F bei 10 eV. Um mit kT diesen Wert zu erreichen, muß $T = 10^5$ K betragen. Bei Zimmertemperatur gilt also $kT \ll \varepsilon_F$, die Leitungselektronen bilden ein stark *entartetes Elektronengas,* d. h. sie zeigen ein von der Boltzmann-Verteilung stark abweichendes Verhalten.

4.10.8. Fermi-Gas

Ein System freier Fermionen nennt man ein Fermi-Gas. Dessen thermodynamische Eigenschaften sollen erläutert werden. Bei Elektronen ist ein Zustand durch die Quantenzahlen Impuls $\boldsymbol{p}$ und Spineinstellung m_S gekennzeichnet. Die zugehörige Energie ist gleich der kinetischen Energie $\varepsilon = \boldsymbol{p}^2/2m$. Wegen der zwei möglichen Spineinstellungen ist in die Zustandsdichte (vgl. 4.10.2.) $g = 2$ einzusetzen $D(\varepsilon) = 4\pi V(2m/\hbar^2)^{3/2}\sqrt{\varepsilon}$. *Teilchenzahl* und *innere Energie* hängen über $N = \int_0^\infty d\varepsilon D(\varepsilon)\, n(\varepsilon, T)$, $U = \int_0^\infty d\varepsilon D(\varepsilon) \varepsilon n(\varepsilon, T)$ mit der Verteilungsfunktion zusammen. Da die mittlere Teilchenzahl N vorgegeben ist, ist die erste Gleichung als Bestimmungsgleichung für das chemische Potential anzusehen.

4.10.8.1. Grundzustand. Bei $T = 0$ gilt $n(\varepsilon, 0) = 0$ bzw. 1 für $\varepsilon > \varepsilon_F$ bzw. $\varepsilon < \varepsilon_F$. Die Fermi-Energie $\varepsilon_F = p_F^2/2m$ trennt die besetzten von den unbesetzten Zuständen. Im Impulsraum (Abb. 4.10.8–1) sind also die Zustände innerhalb einer Kugel vom Radius des *Fermi-Impulses* p_F, innerhalb der sogenannten *Fermi-Kugel,* besetzt. Der Radius der Kugel $p_F = \hbar(3\pi^2 N/V)^{1/3}$ ist durch die Teilchenzahl bzw. die Teilchendichte $\varrho = N/V$ bestimmt. Mit diesem Wert für p_F bzw. ε_F kann man die Zustandsdichte auch als $D(\varepsilon) = (3/2)\left(\sqrt{\varepsilon}/\varepsilon_F^{3/2}\right)N$ schreiben.

Die *Grundzustandsenergie* ist die Summe der kinetischen Energien aller besetzten Zustände $U_0 = N(3/5)\varepsilon_F$. Sie ist proportional zu $\varrho^{2/3}$. Durch Ableitung nach dem Volumen ergibt sich der *Nullpunktsdruck* des Fermi-Gases zu $p_0 = (2/3)U_0/V \sim \varrho^{5/3}$.

4.10.8.2. Entartungsbereich $kT \ll \varepsilon_F$. Für *endliche Temperaturen* wächst die innere Energie an, weil Teilchen in über der Fermi-Energie liegende Zustände angeregt werden. Abb. 4.10.7–1 zeigt, daß die Anzahl der angeregten Teilchen proportional zu kT/ε_F ist. Da die Anregungsenergie im Mittel proportional zu kT ist, wird der Zusatzterm in der inneren Energie proportional zu T^2 sein. Die Auswertung der Integrale für N und U liefert

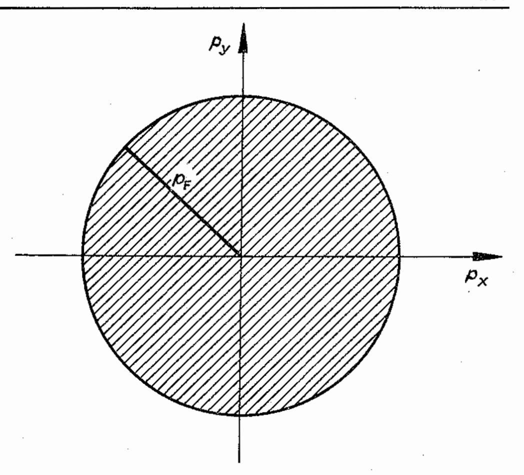


Abb. 4.10.8–1 Fermi-Kugel

$$\mu(T) = \varepsilon_F[1 - (\pi^2/12)(kT/\varepsilon_F)^2 + \ldots],$$
$$U(T, V) = U_0[1 + (5\pi^2/12)(kT/\varepsilon_F)^2 + \ldots].$$

Daraus folgt für den *Elektronenanteil der Molwärme* $C_{El} = (3/2)Nk(\pi^2/3)(kT/\varepsilon_F)$. Dieser Ausdruck ist klein gegen den klassischen Beitrag $3R/2$, wenn $kT \ll \varepsilon_F$ betrachtet wird. Da der Gitteranteil der Molwärme bei tiefen Temperaturen proportional zu T^3 ist, dominiert bei tiefen Temperaturen der Elektronenanteil in der Molwärme $C_{El} + C_G = \gamma T + \alpha T^3$. Unter Verwendung von $D(\varepsilon)$ kann man

$$C_{El} = k(\pi^2/3)\, D(\varepsilon_F)\, kT$$

schreiben. Die Molwärme ist proportional zur Zustandsdichte an der Fermi-Energie, weil nur Elektronen in einem Bereich kT um ε_F angeregt werden können. Mit dieser Formel kann man die Molwärme auch berechnen, wenn Modifikationen der Zustandsdichte durch das Kristallpotential berücksichtigt werden müssen (vgl. 6.3.).

4.11. Quantenelektronik und nichtlineare Optik

4.11.1. Quantenelektronik

Die Quantenelektronik ist ein modernes Forschungsgebiet der Physik, das sich mit der Untersuchung von physikalischen Methoden und der technischen Realisierung von Geräten zur Erzeugung, Verstärkung und Frequenzwandlung kohärenter, elektromagnetischer Signale beschäftigt. Quantenelektronische Verfahren werden gegenwärtig bis in den optischen Frequenzbereich (etwa 10^{16} Hz) angewendet. Es gibt aber schon Vorschläge zur Erzeugung und Verstärkung von kohärenter Röntgen- und γ-Strahlung.

Quantenelektronische Strahlungsquellen zur Erzeugung kohärenter, amplitudenstabiler Strahlung im optischen Spektralbereich werden als *Laseroszillatoren,* Anordnungen zur Verstärkung solcher Strahlung als *Laserverstärker* bezeichnet. Entsprechende Geräte für den Mikrowellenbereich tragen die Bezeichnung *Maser.*

Laser und *Maser* sind Abkürzungen der englischen Bezeichnung *l*ight (bzw. *m*icrowave) *a*mplification by *s*timulated *e*mission of *r*adiation (Licht-[bzw. Mikrowellen-]Verstärkung durch stimulierte Strahlungsemission).

4.11.1.1. Spontane und stimulierte Emission sowie Absorption von Strahlung. Ein Atom besitzt unendlich viele stationäre Zustände mit bestimmter Energie E_i ($i = 1, 2, 3 \ldots$), die sich mit Hilfe quantenmechanischer Methoden, z. B. durch Lösung der Schrödinger-Gleichung, berechnen lassen (vgl. 4.1.5.). Die angeregten Zustände besitzen nur dann eine unendlich lange Lebensdauer und eine genau bestimmte Energie, solange die Wechselwirkung der Atome mit dem Strahlungsfeld nicht berücksichtigt wird (vgl. 4.5.2.).

Aus Gründen der Einfachheit sollen nur die beiden niedrigsten Energieniveaus E_1 und E_2 betrachtet werden. Durch Absorption eines Photons der Energie $h\nu_{21} = E_2 - E_1$ kann ein Atom, das sich im Zustand 1 mit der Energie E_1 befindet, in den angeregten Zustand 2 übergeführt werden. Umgekehrt wird beim Übergang $2 \rightarrow 1$ ein Photon der Energie $h\nu_{21}$ emittiert.

Befindet sich das Atom zum Zeitpunkt t_0 im angeregten Zustand 2 und wirkt keine Strahlung für Zeiten $t \geqq t_0$ auf das Atom ein, so kann dieses im Zustand 2 verbleiben oder verbunden mit der Emission eines Photons der Energie $h\nu_{21}$ in den Zustand 1 mit der Energie E_1 übergehen. Man bezeichnet diesen Vorgang als *spontane Emission*, weil der Übergang $2 \rightarrow 1$ ohne Fremdeinwirkung, wie Zusammenstöße mit anderen Atomen oder Wechselwirkung mit äußerer Strahlung, vonstatten geht. Spontane Emission tritt auch dann auf, wenn sich das Atom isoliert im Vakuum befindet. Infolge der spontanen Emission sind die angeregten Atomzustände instabil.

Die Aussendung des Photons hat zufälligen Charakter. Der Zeitpunkt der Emission kann nicht vorhergesagt werden. Angebbar ist nur die Wahrscheinlichkeit dafür, daß innerhalb eines bestimmten Zeitintervalles eine Emission stattfindet, wenn zu Beginn desselben das Atom im angeregten Zustand war. Die Wahrscheinlichkeit, daß im Verlauf von einer Sekunde das Atom unter Aussendung eines Photons vom Zustand 2 in den Zustand 1 übergeht, wird mit A_{21} bezeichnet. *A_{21} ist die Einsteinsche Übergangswahrscheinlichkeit für spontane Emission.*

Die Übergänge $1 \rightarrow 2$ bzw. $2 \rightarrow 1$ können auch durch Einwirkung äußerer Strahlung verbunden mit Absorption bzw. Emission eines Photons erzwungen werden. Die Wahrscheinlichkeit dafür, daß innerhalb eines bestimmten Zeitintervalles ein solcher Übergang vollzogen wird, ist der Stärke der einfallenden Strahlung proportional. Ein Maß für die Stärke ist die Energiedichte des elektromagnetischen Feldes je Frequenzeinheit $\varrho_\nu(\nu_{21})$ am Ort des Atomes. Die Wahrscheinlichkeit, daß im Verlauf von einer Sekunde das Atom verbunden mit der Absorption eines Photons aus dem Grundzustand 1 in den angeregten Zustand 2 übergeht, ist gegeben durch $B_{12} \cdot \varrho_\nu(\nu_{21})$. Der Proportionalitätsfaktor B_{12} wird als *Einstein-Koeffizient für Absorption* bezeichnet. Zum anderen kann elektromagnetische Strahlung ein Atom, das sich im angeregten Zustand befindet, zur Emission eines Photons stimulieren. Dieser Vorgang, bedingt durch äußere Strahlung, wird als *stimulierte Emission* bezeichnet. Die Wahrscheinlichkeit, daß ein angeregtes Atom im Verlauf einer Sekunde ein Photon durch stimulierte Emission aussendet und dabei in den Grundzustand übergeht, ist $B_{21} \cdot \varrho_\nu(\nu_{21})$. B_{21} ist der *Einstein-Koeffizient für stimulierte Emission.* Für Übergänge zwischen nichtentarteten Atomzuständen gilt $B_{21} = B_{12}$.

Frequenz, Ausbreitungsrichtung, Kohärenzzustand und Polarisationsrichtung der durch stimulierte Emission erzeugten Strahlung stimmen genau mit den entsprechenden Größen der Erregerstrahlung überein. Wird demgemäß Strahlung mit der Frequenz ν_{21} in ein Material bestehend aus vielen Atomen mit der atomaren Übergangsfrequenz ν_{21} geschickt, so ändert sich durch Absorption bzw. stimulierte Emission nur die Anzahl der Photonen im Strahlungsfeld (die Intensität der Strahlung), die anderen Eigenschaften des Strahlungsfeldes bleiben unverändert.

4.11.1.2. Strahlungsverstärkung. Durchläuft Strahlung der Frequenz ν ein Medium bestehend aus Atomen mit einer Übergangsfrequenz $\nu_{21} \approx \nu$, so wird deren Strahlungsleistung im Medium durch Absorption geschwächt und durch stimulierte Emission verstärkt (vgl. 4.11.1.1.). Ist die Anzahl N_2 der angeregten Atome größer als die Anzahl N_1 der nichtangeregten, dann ist die Anzahl der je Zeiteinheit stimuliert emittierten Photonen größer als der je Zeiteinheit absorbierten. Die Strahlung wird verstärkt.

Eine Schwächung der Strahlungsleistung tritt entsprechend ein, wenn $N_2 < N_1$ ist. Da im thermischen Gleichgewicht gemäß der Boltzmann-Verteilung (vgl. 4.10.4.) immer $N_2 < N_1$ gilt, muß, um eine Verstärkung der Strahlungsleistung im Medium zu erreichen, ein vom thermischen Gleichgewicht abweichender Zustand eingestellt werden. Dieser ist dadurch charakterisiert, daß die sogenannte *Besetzungszahlinversion* $(N_2 - N_1)$ größer als Null ist. Das Verhältnis der Strahlungsleistung hinter dem verstärkenden Material $S(L, \nu)$ zur Strahlungsleistung vor Eintritt in dasselbe $S(0, \nu)$ ist gegeben durch:

$$\frac{S(L,\nu)}{S(0,\nu)} = \exp\left[\frac{N_2 - N_1}{V} B_{21} \frac{h\nu_{21}}{c} g(\nu - \nu_{21}) L\right]$$

(B_{21} = Einstein-Koeffizient für stimulierte Emission, $[N_2 - N_1]$ = Besetzungszahlinversion im Material, V = Volumen des durchstrahlenden Mediums, L = Länge des verstärkenden Materiales). $g(\nu - \nu_{21})$ bezeichnet die *Linienformfunktion.* Diese Funktion berücksichtigt die Tatsache, daß die angeregten Niveaus der Atome nicht unendlich schmal, sondern infolge der Wechselwirkung der Atome mit ihrer Umgebung verbreitert sind. Das führt dazu, daß auch Strahlung mit Frequenzen, die nicht genau mit der Resonanzfrequenz ν_{21} übereinstimmen, verstärkt werden. $g(\nu - \nu_{21})$ stellt den Frequenzgang des Strahlungsverstärkers dar. Analog zur Elektronik werden Materialien, die fähig sind, Strahlung zu verstärken, als *aktive Materialien* bezeichnet. Das Eigenrauschen eines solchen quantenelektronischen Verstärkers setzt sich zusammen aus einem thermischen Rauschanteil und einem Anteil, der durch die im aktiven Medium stattfindenden spontanen Emissionsprozesse verursacht wird.

4.11.1.3. Erzeugung der Besetzungszahlinversion. *Inversions-* oder *Pumpmechanismen* gestatten es, im aktiven Material eine positive Besetzungszahlinversion $(N_2 - N_1) > 0$ zu erzeugen. Durch einen Pumpprozeß, z. B. durch Pumpstrahlung oder unelastische Stöße, vom unteren Niveau E_1 zum oberen Energieniveau E_2 eines *Zweiniveausystems* ist keine Besetzungszahlinversion erreichbar. Bei sehr starkem Pumpen kann höchstens eine *Sättigung des Überganges* mit $N_2 = N_1$ erzielt werden.

Eine Sonderstellung nimmt hierbei der *Molekularstrahlmaser* (vgl. 4.11.1.4.) ein. Das aktive Medium im Molekularstrahlmaser stellt ein Zweiniveausystem dar. Die Besetzungszahlinversion wird in diesem Fall nicht durch zusätzliches Pumpen, sondern durch eine spezielle Sortiervorrichtung zur Trennung von angeregten und nichtangeregten Molekülen erreicht.

Im Unterschied zum Zweiniveausystem kann durch starkes

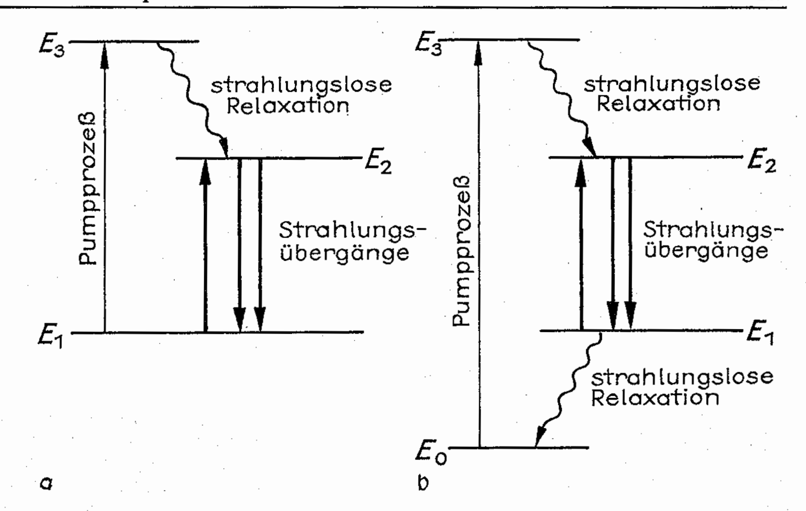


Abb. 4.11.1–1 (a) Drei- und (b) Vierniveausystem

Pumpen in *Drei-* oder *Vierniveausystemen* eine positive Besetzungszahlinversion hergestellt werden (Abb. 4.11.1–1). Werden entsprechende Materialien in Lasern eingesetzt, so spricht man vom Drei- bzw. Vierniveaulaser. Beim Dreiniveauschema werden Atome des aktiven Materiales durch einen Pumpprozeß vom Niveau 1 nach Niveau 3 übergeführt. Durch sehr schnelle strahlungslose Übergänge *(Relaxation)* gelangen diese in das obere Niveau 2. Wegen der schnellen Relaxationsprozesse $3 \rightarrow 2$ ist die Besetzungszahl N_3 stets kleiner als die des Grundniveaus. Der Übergang 1 nach 3 kann deshalb nicht gesättigt werden, und es ist prinzipiell möglich, alle Atome aus dem Energieniveau E_1 in den angeregten Zustand mit der Energie E_2 zu pumpen.

Das Vierniveausystem hat gegenüber dem Dreiniveausystem den Vorteil, daß das untere Niveau 1 durch Relaxationsprozesse schnell entleert wird. Eine positive Besetzungszahlinversion zwischen den Niveaus 2 und 1 ist deshalb schon bei Besetzungszahlen N_2 erreichbar, die viel kleiner sind als die Gesamtzahl der Atome im aktiven Material, weil die spontane Absorption nur als Konkurrenz auftritt, wenn das Niveau 1 besetzt ist. Im Dreiniveausystem tritt Inversion erst dann ein, wenn N_2 größer ist als die Hälfte der Gesamtzahl der aktiven Atome. Dreiniveausysteme werden auch bei aktiven Masermaterialien zur Herstellung der Besetzungsinversion angewendet (Abb. 4.11.1–2). Ein effektiver Pumpprozeß bei einem Dreiniveaumaser ist wegen der geringen Energiedifferenzen nur bei sehr tiefen Temperaturen möglich. Um im thermischen Gleichgewicht merkliche Besetzungsunterschiede der drei Niveaus zu erreichen, werden aktive Masermaterialien auf Temperaturen des flüssigen Heliums abgekühlt. Strahlungsübergänge zwischen den drei Niveaupaaren 1-3, 3-2 und 2-1 sind erlaubt. Durch Einstrahlung einer Pumpwelle der Frequenz ν_{31} wird die Besetzungszahl N_3 erhöht und N_1 entsprechend verringert. Bei hinreichend starker Pumpstrahlung wird Gleichbesetzung $N_3 = N_1$ (gestrichelte Linie in Abb. 4.11.1–2) erzielt.

4.11.1.4. Aufbau, Wirkungsweise und Anwendung von Masern. Der prinzipielle Aufbau eines *Molekularstrahlmaser-Verstärkers* ist in Abb. 4.11.1–3 dargestellt. Einer der ersten dieser Masertypen war der Ammoniakstrahlmaser.

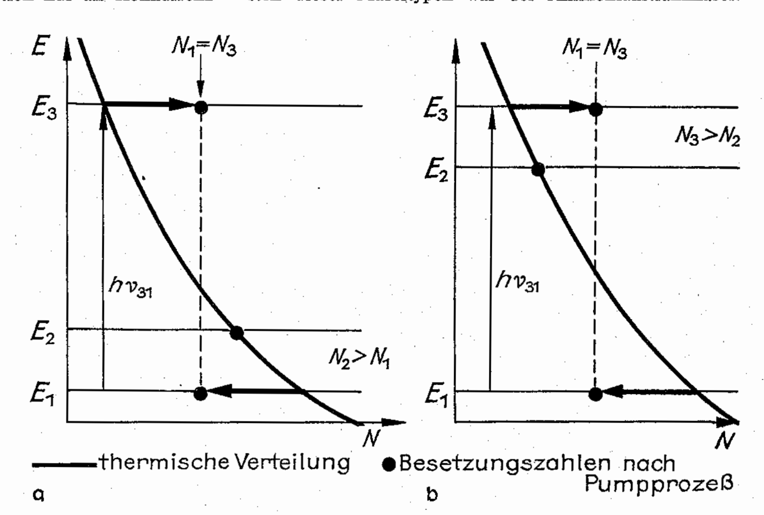


Abb. 4.11.1–2 Erzeugung der Besetzungszahlinversion beim Dreiniveaumaser, (a) Besetzungsinversion zwischen den Niveaus 2 und 1, (b) zwischen den Niveaus 3 und 2

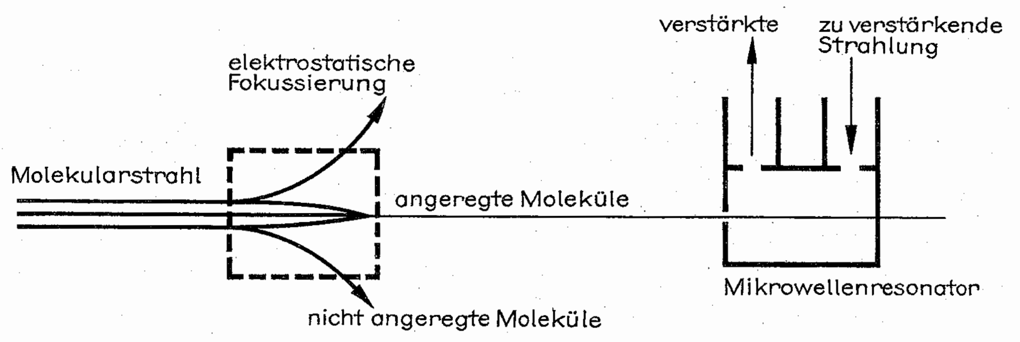


Abb. 4.11.1–3 Molekularstrahlmaser-Verstärker

Ausgenutzt für den Maserübergang wird eine bestimmte Schwingung des Ammoniakmoleküles mit einer Emissionswellenlänge von $\lambda = 12{,}7$ mm, bei der das Stickstoffatom durch die von den drei Wasserstoffatomen aufgespannte Ebene hindurchschwingt. Die Ammoniakmoleküle passieren ein Gebiet, in dem durch eine Spezialvorrichtung ein inhomogenes elektrostatisches Feld bestimmter Form erzeugt wird. In diesem Feld erfährt ein elektrisches Dipolmoment eine Kraftwirkung (vgl. 2.4.1.2.). Da die angeregten Moleküle dipollos sind, während die Moleküle im Grundzustand ein von Null verschiedenes Dipolmoment besitzen, werden die nicht angeregten Moleküle seitlich abgelenkt, die angeregten laufen unbeeinflußt weiter. Diese Selektion der Moleküle ist insofern effektiv, weil wegen der geringen Energiedifferenz zwischen den Maserniveaus ($E_2 - E_1 \ll kT$) beide im Ausgangsstrahl etwa gleich besetzt sind. Der Molekülstrahl durchläuft dann einen Mikrowellenhohlraumresonator mit Eigenfrequenzen, die nahe der Frequenz der Ammoniakschwingung liegen. Mikrowellenresonatoren sind Hohlräume mit Metallwänden, deren Ausmaße vergleichbar mit der Wellenlänge der Strahlung sind. Da nicht die gesamte Strahlung den Hohlraumresonator verläßt, sondern ein Teil davon durch Reflexion im Resonator verbleibt und erneut durch stimulierte Emission der angeregten Moleküle verstärkt wird, kann es bei genügend hoher Verstärkung zur Selbsterregung des Systemes kommen. Man bezeichnet eine solche Anordnung, bei der durch Rückkopplung eine Selbsterregung eintritt, als *Maseroszillator.* Ein typischer *Festkörpermaser* ist der Rubinmaser, bei dem das aktive Material ein Dreiniveausystem darstellt. Als Maserniveaus dienen *Zeeman-Niveaus,* die durch Aufspaltung bestimmter Energieniveaus in einem magnetischen Gleichfeld entstehen (vgl. 4.5.3.). Festkörpermaser werden durch Einstrahlung von Mikrowellen mit der Frequenz $\nu = \nu_{31}$ gepumpt. Um den Pumpmechanismus entsprechend dem Dreiniveauschema zu ermöglichen, müssen Festkörpermaser bei Temperaturen des flüssigen Heliums betrieben werden.

Maserverstärker zeichnen sich durch ein äußerst geringes Rauschen aus und werden deshalb mit Erfolg zur Verstärkung extrem schwacher Signale benutzt. Bewährt haben sie sich bei der Satellitennachrichtenübertragung.

Maseroszillatoren haben eine sehr hohe Frequenzstabilität. Das gestattet ihren Einsatz als *Frequenznormale (Atomuhren).* Weiterhin dienen sie als Generatoren für mm- und cm-Wellen. Maseroszillatoren können sowohl in der Amplitude als auch in der Frequenz moduliert werden. Eine Frequenzmodulation erreicht man über die Variation des Niveauabstandes durch Änderung der magnetischen Feldstärke. Eine Amplitudenmodulation wird durch gezielte Veränderungen der Besetzungszahlinversion erreicht.

4.11.1.5. Laser. Eigenschaften der Laserstrahlung. Die im Laseroszillator erzeugte Strahlung zeichnet sich durch hohe *Monochromasie* aus. Die kleinste erreichbare Linienbreite $\Delta\nu_L$ wird wie bei einem Radiosender durch Rauschprozesse bestimmt. Im optischen Spektralbereich wird das Rauschen im wesentlichen durch die spontane Emission verursacht. Mit frequenzstabilisierten Gaslasern konnten minimale Werte für die relative Linienbreite $\Delta\nu_L/\nu_L \approx 10^{-14}$ erzielt werden. Die sehr geringe Linienbreite der Laserstrahlung bewirkt eine hohe *zeitliche Kohärenz* derselben. Die entsprechende Kohärenzlänge $c/\Delta\nu_L$ kann direkt mit *Michelson-Interferometern* gemessen werden. Die Kohärenzeigenschaften von Laserlicht und Lichtthermischer Quellen, das durch optische Filter auf die gleiche Linienbreite $\Delta\nu_L$ eingeengt wird, stimmen nicht vollständig überein. Ursache dafür ist die unterschiedliche Amplitudenverteilung der vom Laser abgestrahlten Lichtwellen und der thermisch emittierten Wellen. Man spricht davon, daß die Laserstrahlung, die durch eine nahezu konstante Feldstärkeamplitude gekennzeichnet ist, in höherer Ordnung kohärent ist als die Strahlung thermischer Lichtquellen.

Laserstrahlung zeigt eine hohe *räumliche Kohärenz.* Da nur Lichtwellen mit bestimmter Phase über den gesamten Strahlquerschnitt am Ort der Resonatorspiegel Eigenschwingungen des elektromagnetischen Feldes im Resonator bilden können, sind die Lichtwellen, die aus dem Resonator ausgekoppelt werden, durch feste Phasenbeziehungen über den Querschnitt des Lichtbündels gekennzeichnet.

Im Laseroszillator werden merklich nur die Wellen verstärkt, die in Richtung Resonatorachse verlaufen. Die minimale *Divergenz* der Strahlung ist durch die Beugung gegeben. Die beugungsbegrenzte Divergenz führt zu einem Divergenzwinkel $\Theta_{min} \approx \lambda/D$ (λ = Wellenlänge der Strahlung, D = Bündeldurchmesser). Mit $\lambda \approx 5 \cdot 10^{-4}$ mm und $D \approx 5$ mm erhält man $\Theta_{min} \approx 0{,}006°$.

Wichtige Lasertypen. *Festkörperlaser.* Die aktiven Materialien eines Festkörperlasers haben meist die Form langer, dünner Zylinder (Abb. 4.11.1–4). Festkörperlaser werden optisch mit Hilfe von Gasentladungslampen gepumpt. Das Pumplicht wird mittels Reflektor in das aktive Medium konzentriert. Wichtige Vertreter der Festkörperlaser sind die Rubinlaser (Dreiniveausystem) und die Neodymlaser (Vierniveausystem). Festkörperlaser arbeiten in den meisten Fällen im Impulsbetrieb. Die Pumpdauer beträgt etwa 500 μs. Bei Verwendung einer sogenannten Güteschaltung

gelingt es, während der Pumpdauer den Laser unterhalb der Laserschwelle zu halten und dadurch eine große Besetzungsinversion aufzubauen. Man erreicht das durch spezielle Drehspiegel, elektrooptische Schalter oder zusätzliche Absorber im Resonator. Als Ergebnis erhält man sogenannte *Riesenimpulse* sehr großer Leistung (bis etwa 100 MW) und kurzer Dauer. Durch *Modensynchronisation* können noch kürzere Impulse erzielt werden. Modensynchronisation wird erreicht durch aktive Bauelemente im Resonator, die die Resonatorverluste periodisch mit der Modenabstandsfrequenz $c/2L$ variieren, oder durch Absorber mit sehr kurzer Lebensdauer im angeregten Niveau. Man erhält eine Folge von sehr kurzen Impulsen großer Leistung (etwa 1 GW).

Gaslaser. Mit den bekannten Gaslasern läßt sich eine Laserstrahlung vom mittleren IR-Bereich bis in den UV-Spektralbereich erzeugen. Bis auf wenige Ausnahmen werden die Gaslaser elektrisch gepumpt. Bei Niederdruckgaslasern ($p = 10^2$ bis 10^3 Pa) brennt die Gasentladung kontinuierlich. Bei Hochdruckgaslasern ($p \lesssim 10^5$ Pa) wird das Gasgemisch transversal zur Resonatorachse mit kurzen elektrischen Impulsen angeregt. Die Energie zur Erzeugung der Besetzungsinversion wird durch unelastische Atomstöße auf die aktiven Atome übertragen. Gaslaser arbeiten im kontinuierlichen und im Impulsbetrieb. Auch bei ihnen ist Güteschaltung und Modensynchronisation zur Erzeugung kurzer Impulse möglich.

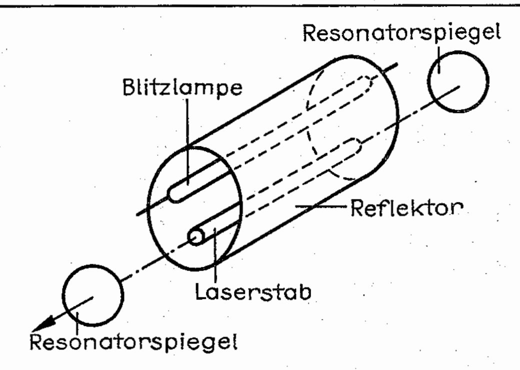


Abb. 4.11.1–4 Aufbau eines Festkörperlasers

Farbstofflaser verwenden als aktive Materialien Lösungen organischer Farbstoffe mit großer *Fluoreszenzquantenausbeute.* Farbstofflaser sind Vierniveaulaser. Der Laserüber-

Tab. 4.11.1–5 Typische Laserparameter

Typ	Wellenlänge	Pumpart	Betriebsart	Impulsdauer	Strahlungsleistung
Festkörperlaser					
Rubinlaser	694 µm	optisch, Blitzlampe	gütegeschaltet	20 ns	10...50 MW
			modensynchronisiert	20 ps	1...10 GW
Neodymlaser	1,06 µm	optisch, Blitzlampe	gütegeschaltet	10...100 ns	10...100 MW
			modensynchronisiert	10 ps	1...10 GW
Gaslaser					
He-Ne-Laser	633 nm	elektrisch, kontinuierlich	kontinuierlich		1...10 mW
	nahes IR		modensynchronisiert	100 ps	einige kW
Argonlaser	458...515 nm	elektrisch, kontinuierlich	kontinuierlich		0,5...20 W
			modensynchronisiert	150 ps	einige kW
Kupferdampflaser	511 nm und 578 nm	elektrisch, gepulst	gepulst	100...500 ns	100...500 kW
Stickstofflaser	337 nm	elektrisch, gepulst	gepulst	0,1...10 ns	100 kW...1 MW
Kryptonfluoridlaser	248 nm	elektrisch, gepulst	gepulst	500 ns	100 MW
Iodlaser	1,3 µm	optisch, Blitzlampe	Pulsformung	10 ns	< 100 GW
Niederdruck-CO_2-Laser	10,6 µm	elektrisch, kontinuierlich	kontinuierlich		1 W...10 kW
			gütegeschaltet		100 kW
Hochdruck-CO_2-Laser	10,6 µm	elektrisch, gepulst	gepulst	200 ns...20 µs	1 MW...10 GW
Farbstofflaser	350 nm	optisch, Blitzlampe	kontinuierlich		0,5...10 W
	1 000 nm	Argonlaser	modensynchronisiert	< 0,1...10 ps	100 W...1 kW
Halbleiterlaser					
Galliumarsenidlaser	820 nm	elektrisch, kontinuierlich	kontinuierlich		100 mW
	880 nm	gepulst	gepulst	100 ns	20 W
chemische Laser					
HF-Laser	IR	chemische Reaktionen, diskontinuierlich	gepulst	50 ns	< 100 GW

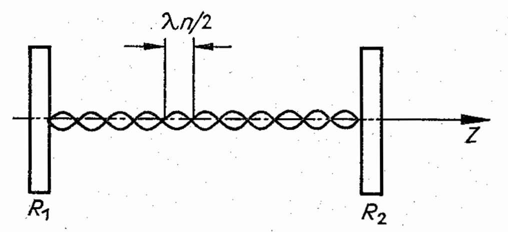


Abb. 4.11.1–6 stehende Welle im planparallelen Resonator

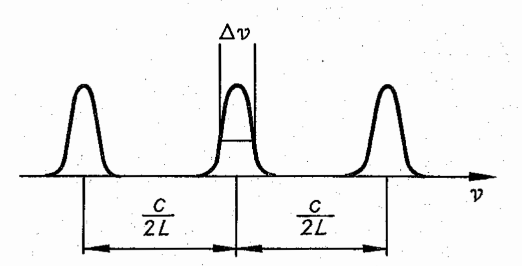


Abb. 4.11.1–7 Eigenfrequenzen des Resonators

gang erfolgt zwischen Elektronenschwingungszuständen des ersten angeregten Elektronenzustandes und des Elektronengrundzustandes von organischen Molekülen. Während die bekannten Festkörper- und Gaslaser in einem engen Wellenlängenintervall (bestimmt durch die relativ schmalen Linien des aktiven Materials) strahlen, kann die Emissionswellenlänge beim Farbstofflaser durch wellenlängenselektive Elemente im Resonator über den breiten Wellenlängenbereich der Farbstoffbanden verändert werden. Mit Hilfe von Farbstofflasern kann durch Einsatz verschiedener Farbstofflösungen Laserlicht mit Wellenlängen zwischen 350 und 1000 nm erzielt werden. Aktive Farbstofflasermaterialien werden optisch mittels Blitzlampen oder mit Laserstrahlungsquellen gepumpt. Wegen der sehr großen Bandbreite des Verstärkermediums von Farbstofflasern ist es möglich, durch Modensynchronisation Lichtimpulse mit einer Dauer unterhalb 1 ps zu erzeugen.

Halbleiterlaser nutzen zur Laseremission optisch erlaubte Übergänge zwischen Leitfähigkeitsband und Valenzband aus (vgl. 6.3.2.). Mittels verschiedener Methoden, wie Anlegen äußerer elektrischer Felder oder *Ladungsträgerinjektion*, kann ein Elektronenüberschuß im Leitfähigkeitsband erzeugt werden. Infolge sogenannter *Intrabandübergänge* geben die Elektronen einen Teil ihrer Energie innerhalb sehr kurzer Zeit (≈ ps) strahlungslos an das Kristallgitter ab. Nach der Energieabgabe gelangen die Elektronen aus angeregten Zuständen des Leitungsbandes zu Zuständen an der unteren Grenze desselben. Es besteht dann eine Besetzungsinversion zwischen den niedrigsten Zuständen des Leitungsbandes und den höchsten des Valenzbandes.

Die Frequenz der Halbleiterlaser läßt sich durch Druck oder äußere Felder verändern.

Chemische Laser erzeugen die Besetzungsinversion durch schnelle chemische Reaktionen. Zur Laseremission werden *infrarotaktive* Schwingungsübergänge von Molekülen ausgenutzt. Die miteinander reagierenden Gase werden getrennt in einen gemeinsamen Reaktionsraum geblasen. Infolge der dort stattfindenden Reaktion werden bei einer Molekülsorte angeregte Schwingungsniveaus sehr stark bevölkert. Zwischen den angeregten Schwingungsniveaus und den energetisch tieferliegenden besteht dann eine Besetzungsinversion. Tab. 4.11.1–5 gibt einen Überblick über typische Parameter der verschiedenen Laserarten.

Laserresonatoren. Der einfachste Resonator für Lichtwellen sind parallel zueinander justierte ebene Spiegel. Das elektromagnetische Feld in einem solchen Resonator besteht aus sehr vielen zwischen den Spiegeln hin- und herlaufenden ebenen Wellen. Dies führt für Licht der Wellenlängen λ_n mit $n \cdot \lambda_n/2 = L$ ($n = 1, 2, 3 \ldots$, L = Abstand zwischen den Spiegeln) zu stehenden Wellen mit stationären Feldverteilungen im Resonator (Abb. 4.11.1–6). Sie werden als *Eigenschwingungen* oder *Moden* des Resonators bezeichnet. Der Frequenzabstand zwischen den Moden benachbarter Modenindizes beträgt $c/2L$ (c = Lichtgeschwindigkeit). Er ist so klein, daß sich mehrere Frequenzen innerhalb der durch g bestimmten Linienbreite befinden (Abb. 4.11.1–7). Die Lichtwellenlängen liegen im Bereich von $4 \cdot 10^{-7}$ m bis 10^{-6} m und L ist 0,1 bis 1 m, demzufolge sind die *Modenindizes* n sehr große Zahlen ($n \approx 10^5$ bis $2{,}5 \cdot 10^6$). Da ein Teil der Strahlung durch die Spiegel ausgekoppelt wird und auch andere Verluste für die Lichtwellen im Resonator auftreten, werden die einmal im Resonator angeregten Moden gedämpft. Die Dämpfung führt zu einer Modenlinienbreite $\Delta\nu$. $\Delta\nu$ hängt wesentlich vom Spiegelabstand L und den Reflexionsvermögen R_1 und R_2 der beiden Spiegel ab. Wenn die Lichtwellen im Resonator senkrecht zur Resonatorachse z nicht unendlich weit ausgedehnt sind, ergeben sich Feldverteilungen, die in komplizierter Weise von allen drei Raumkoordinaten abhängen.

Laseroszillatoren bestehen aus einem Laserverstärker (vgl. 4.11.1.2.) und einem optischen Resonator. Das aktive Medium bewirkt durch den Prozeß der Lichtverstärkung eine

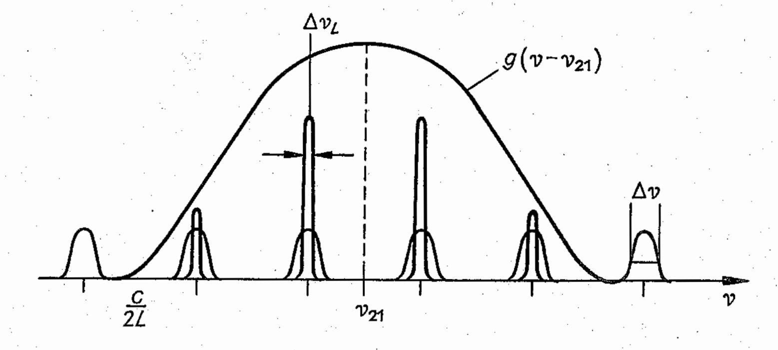


Abb. 4.11.1–8 Verstärkung und Entdämpfung der Resonatormoden durch das aktive Medium

Entdämpfung der Eigenschwingungen des elektromagnetischen Feldes im Resonator. Die Energieverluste der hin- und herlaufenden Wellen, die hauptsächlich dadurch entstehen, daß ein Teil der Strahlung an den Resonatorspiegeln den Resonator verläßt, werden beim Durchlauf dieser Wellen durch das aktive Material durch stimulierte Emission ausgeglichen. Die Entdämpfung führt weiterhin zu einer Verringerung der Modenlinienbreite $\Delta\nu$. Verstärkt werden aber nur die Eigenschwingungen des Resonators, deren Frequenzen innerhalb der Verstärkerkurve $g\,(\nu-\nu_{21})$ (vgl. 4.11.1.2.) des aktiven Medium liegen (Abb. 4.11.1–8). Eine *Selbsterregung* des Laseroszillators setzt für bestimmte Moden dann ein, wenn bei Hin- und Rücklauf der Strahlung durch das verstärkende Material die bei Hin- und Rücklauf an den beiden Spiegeln durch Auskopplung entstandenen Verluste überkompensiert werden, wenn also gilt

$$R_1 \cdot R_2 \cdot \exp\left[2 \cdot \frac{N_2 - N_1}{V} B_{21} \frac{h\nu_{21}}{c} \cdot g\,(\nu \cdot \nu_{21})\,L\right] \geq 1$$

(vgl. 4.11.1.2.).

Diese Beziehung wird als *Schwellbedingung* des Lasers bezeichnet. Oberhalb der Laserschwelle arbeitet der Laseroszillator als Sender von kohärenter Strahlung. Das im aktiven Material zufällig in Resonatorachsenrichtung (vgl. Abb. 4.11.1–4, z-Achse) spontan emittierte Licht, das noch in Frequenz- und Polarisationsrichtung mit Resonatoreigenschwingungen innerhalb der Verstärkerkurve übereinstimmt, wird beim Hin- und Herlauf zwischen den Resonatorspiegeln im aktiven Medium verstärkt. Der Laser ist ein rückgekoppelter Verstärker für die spontane Strahlung.

Aus der Schwellbedingung kann die zur Selbsterregung nötige Besetzungsinversion und damit die erforderliche Stärke des Pumpprozesses berechnet werden.

Anwendung von Lasern. Die besonderen Eigenschaften der Laserstrahlung ermöglichen vielfältige Anwendungen von Laserstrahlungsquellen in Wissenschaft und Technik.

Standards für Länge und Frequenz. Stabilisierte Laser gewährleisten eine höhere relative Genauigkeit der Lichtwellenlänge als das bisherige *Längenstandard* auf der Basis der Kryptonlampe. Das Meter ergibt sich dabei als ein Vielfaches der Strahlungswellenlänge. Laser sind auch wegen der großen Kohärenzlängen ($c/\Delta\nu_L > 1$ m) als Eichlichtquellen zur Meterdefinition geeignet. Mit einem durch Iodgas stabilisierten He-Ne-Laser kann die Längeneinheit genauer als 10^{-10} festgelegt werden. Die Kryptonlampe garantiert eine Genauigkeit der Festlegung von 10^{-8}. Die Strahlung des iodstabilisierten He-Ne-Lasers kann auch als *Frequenzstandard* ausgenutzt werden, wobei relative Genauigkeiten von 10^{-13} erreichbar sind. Durch vielfache stufenweise *Frequenzdifferenzmessung* zwischen den Frequenzen geeigneter Laser mittels spezieller Fotoempfänger mit hoher Grenzfrequenz – sogenannter *Punktkontaktdioden* – ist es möglich, die Strahlungsfrequenzen im optischen Spektralbereich direkt mit der Frequenz des bisherigen Frequenzstandards, der *Caesiumuhr*, zu vergleichen.

Durch absolute Messungen von Wellenlänge und Frequenz der Laserstrahlung konnte der Wert für die Lichtgeschwindigkeit mit bisher nicht erreichter Genauigkeit bestimmt werden (vgl. 2.1.1.1.).

Hochauflösende Laserspektroskopie. Wegen der hohen Frequenzstabilität und der sehr geringen Linienbreiten gelingt es, mit Laserstrahlungsquellen sehr kleine Frequenzabstände bis zu einigen 100 Hz zu messen. Hyperfeinstrukturaufspaltungen und Frequenzverschiebungen von Spektrallinien können sehr genau ermittelt werden. Es gelang mit Hilfe der hochauflösenden Laserspektroskopie z. B., die Spektralverschiebung der Elektronenterme zu vermessen, die durch die Masseänderung des Kernes ΔM_K infolge Kernanregung ΔE nach $\Delta M_K = \Delta E/c^2$ verursacht wurden. Der bisher genaueste Wert für die Rydberg-Konstante (vgl. 4.2.2.) wurde mit laserspektroskopischen Verfahren bestimmt.

Durchstimmbare, schmalbandige Laserlichtquellen gestatten eine selektive spektrale Anregung bestimmter angeregter Energieniveaus. Dadurch wird es möglich, *chemische Reaktionen* gezielt zu stimulieren und *Isotopentrennung* auf neue Weise durchzuführen.

Laserkurzzeitspektroskopie. Mit Hilfe der von gütegeschalteten und modensynchronisierten Lasern erzeugten kurzen und leistungsstarken Lichtimpulse können kurzzeitig angeregte Zustände in Festkörpern und Molekülen aufgebaut und durch zeitaufgelöste Messungen der Lichtabsorption oder -emission die Lebensdauer und die Spektren solcher Zustände ermittelt werden. Es werden dabei Zeitauflösungen bis unter 0,1 ps erreicht.

Lasermikrospektralanalyse. Die Fokussierbarkeit von beugungsbegrenzten Laserstrahlbündeln bis in die Dimension von einigen Mikrometern gestattet die gezielte spektroskopische Untersuchung von mikroskopisch kleinen Ausschnitten von Werkstoffoberflächen. Durch die fokussierte Laserstrahlung wird Oberflächenmaterial verdampft und zur Emission der charakteristischen Spektrallinien angeregt.

Materialbearbeitung. Die großen Dauer- und Impulsstrahlungslleistungen sowie die Möglichkeit, die Strahlung auf kleine Flächen zu fokussieren, bieten verschiedene Varianten, den Laser als Strahlungsquelle in Bohr-, Schneid- und Schweißroboter einzubauen. In ähnlicher Weise werden sogenannte *Laserskalpelle* in der Chirurgie verwendet. In der Augenheilkunde wird der Laser bei Netzhautoperationen erfolgreich eingesetzt.

Gesteuerte Kernfusion. Mit Festkörperlasern und nachfolgenden Verstärkerstufen sind Strahlungsintensitäten von $10^{22}\,\mathrm{W/m^2}$ zu erreichen. Die Einwirkung dieser intensiven Strahlung auf spezielle Targets gestattet es, Plasmen sehr hoher Temperatur ($\gtrsim 10^8$ K) zu erzeugen. Daraus resultiert eine Möglichkeit, mit solchen Laseranlagen die gesteuerte Kernfusion zu realisieren.

Längenmeßtechnik. Für Entfernungsmessungen von langen Wegstrecken (10 bis 100 km) auf der Erde, zur genauen Bestimmung von Entfernungen zwischen künstlichen Erdsatelliten und Beobachtungsstationen auf der Erde sowie Entfernungsmessungen Erde – Mond wird die Methode des *Laserradars* verwendet. Ausgenutzt werden dabei die geringe Winkeldivergenz und große Impulsleistung der von einem Festkörperimpulslaser abgegebenen Laserstrahlung. Die Entfernung wird über die Laufzeit der am entfernten Objekt reflektierten Lichtimpulse gemessen. Mittlere terrestrische Entfernungen (100 m bis 10 km) werden durch Phasenmessung eines modulierten Laserstrahles mit Millimetergenauigkeit bestimmt. Zur Messung von Längen unter 10 m werden Laserinterferometer verwendet; sie gestatten Genauigkeiten von 0,01 µm.

Durch Kopplung spektroskopischer Verfahren, wie Resonanzstreuung, Ramanstreuung u. ä., mit dem Laserradar ist es möglich, Luftverschmutzungen in der Atmosphäre über große Entfernungen zu lokalisieren und empfindlich nachzuweisen. Dieses Verfahren hat große Bedeutung für den *Umweltschutz.*

Informationsübertragung. Seit der Entwicklung von Glasfasern sehr geringer Lichtschwächung (einige dB/km), hat die Nachrichtenübertragung mit Hilfe elektromagnetischer Strahlung im optischen Spektralbereich stark an Bedeutung gewonnen. Als Sender und Verstärker werden Halbleiterlaser und Halbleiterlaserverstärker verwendet.

Holografie ist ein kohärent optisches Abbildungsverfahren, bei dem auf der Fotoplatte die vom Objekt beeinflußte Lichtwelle mit einer Vergleichswelle der gleichen Lichtquelle überlagert wird. Dadurch wird nicht nur die Amplitude, sondern auch die Phase der Objektwelle gespeichert. Eine erneute Bestrahlung der entwickelten Fotoplatte, des sogenannten *Hologrammes,* mit der Vergleichswelle läßt ein dreidimensionales Bild des Objektes entstehen. Der wesentliche Vorteil der Laserstrahlung gegenüber der spektral gefilterten Strahlung thermischer Lichtquellen für die Holografie ist die wesentlich höhere Intensität bei gleichem Kohärenzgrad.

4.11.2. Nichtlineare Optik

Unter dem Begriff *nichtlineare Optik* faßt man alle elektromagnetischen Erscheinungen bei hohen Frequenzen ($\nu \gtrsim 10^{11}$ Hz) zusammen, die auf der Nichtlinearität der Materialgleichungen im System der *elektromagnetischen Feldgleichungen,* den *Maxwell-Gleichungen* (vgl. 2.4.), beruhen. Nichtlinearitäten treten bei der Wechselwirkung des Lichtes mit atomaren Systemen in Erscheinung, sobald die Feldstärke der Lichtwellen nicht vernachlässigbar klein gegenüber den inneren Feldern ist, die auf die Elektronen wirken und deren Bindung an das Kerngerüst verursachen. Während es mit den üblichen thermischen Lichtquellen kaum möglich war, hinreichend starke Felder zu erzeugen, erlauben Laser die Erzeugung von Feldern, die starke Nichtlinearitäten und sogar die Ablösung von Elektronen und damit schließlich die Zerstörung von atomaren Systemen bewirken können. Aus diesen Gründen konnte sich die nichtlineare Optik erst nach dem Bau der ersten Laser entwickeln.

4.11.2.1. Nichtlineare Materialgleichungen. Nichtlineare Abhängigkeiten der elektrischen Polarisation von der elektrischen Feldstärke bzw. der Magnetisierung von der magnetischen Feldstärke sind bereits seit dem 19. Jh. bekannt. Dabei handelt es sich einmal um Prozesse, an denen nur Gleichfelder oder Felder mit kleinen Frequenzen beteiligt sind – Beispiele dafür sind *parelektrische* bzw. *paramagnetische* sowie *ferroelektrische* bzw. *ferromagnetische* Erscheinungen (vgl. 6.4.) –, und zum anderen um Prozesse, bei denen die Brechzahl für Lichtwellen durch starke Felder niedriger Frequenz beeinflußt wird, wie dies beispielsweise beim *Pokkels-*, *Kerr-* und *Faraday-Effekt* auftritt.

Bei Verwendung von Lasern können die starken äußeren elektromagnetischen Felder solche Kräfte auf die Elektronenwolke der atomaren Systeme ausüben, daß im Zusammenhang zwischen der Auslenkung der Elektronenwolke aus der Gleichgewichtslage und der Lichtfeldstärke die nichtlinearen Terme nicht mehr zu vernachlässigen sind. Als Folge ergibt sich eine nichtlineare Abhängigkeit der Polarisation P von der Lichtfeldstärke E

$$P = \varkappa_1 E + \varkappa_2 E^2 + \varkappa_3 E^3 + \ldots = P_L + P_{NL},$$

die zu völlig neuen optischen Erscheinungen führt. Eine monochromatische Anregungsfeldstärke $E = E_0 \cdot \cos \omega t$ ergibt z. B.

$$P_L = \varkappa_1 E_0 \cos \omega t,$$

$$P_{NL} = \varkappa_2 E_0^2 \cos^2 \omega t + \varkappa_3 E_0^3 \cos^3 \omega t + \ldots$$

$$= (1/2) \varkappa_2 E_0^2 + (1/2) \varkappa_2 E_0^2 \cos 2\omega t + (3/4) \varkappa_3 E_0^3 \cos \omega t + (1/4) \varkappa_3 E_0^3 \cos 3\omega t + \ldots$$

Gemäß den Potenzen der Feldstärke spricht man von nichtlinearen Polarisationsanteilen zweiter Ordnung, dritter Ordnung usw. und dementsprechend von nichtlinearen Effekten zweiter, dritter und höherer Ordnung. Man erkennt sofort, daß nur im Rahmen der linearen Näherung die Zeitabhängigkeit der Polarisation – und als Folge auch die Zeitabhängigkeit der durch das erregte Material abgestrahlten Wellen – identisch mit der Zeitabhängigkeit des erregenden Feldes ist. Bereits die ersten Glieder des nichtlinearen Polarisationsanteiles zeigen, daß Anteile mit qualitativ völlig anderen Zeitabhängigkeiten auftreten:

a) ein in der Feldstärke quadratischer Anteil bei der Frequenz Null, der eine *Gleichrichtung optischer Signale* und damit das Entstehen einer Gleichspannung an der Probe bewirkt,
b) ein weiterer in der Feldstärke quadratischer Anteil bei der doppelten Frequenz, der die Erzeugug von Licht bei der doppelten Frequenz, die *s*econd *h*armonic generation *(Erzeugung der zweiten Harmonischen),* kurz SHG genannt, verursacht,
c) ein in der Feldstärke kubischer Anteil bei der Grundfrequenz, der eine *Intensitätsabhängigkeit der Brechzahl* und in der Folge intensitätsabhängige Ausbreitungseigenschaften, z. B. die *Selbstfokussierung* eines Lichtbündels in einem Material, bewirkt,
d) ein weiterer in der Feldstärke kubischer Anteil bei der dreifachen Frequenz, der die Erzeugung von Licht bei der dreifachen Frequenz, die *t*hird *h*armonic generation, kurz THG genannt, hervorruft.

Der gleiche Sachverhalt kann anhand der Kennlinie $P(E)$ des Materiales in Analogie zur Betrachtungsweise in der Elektronik verdeutlicht werden (Abb. 4.11.2–1): Bei Einstrahlung einer streng harmonischen Lichtfeldstärke entsteht eine nichtharmonische Polarisationswelle, die man in eine Welle bei der Grundfrequenz und in Oberwellen zerlegen kann.

Das Erscheinungsbild wird noch vielfältiger, wenn man anstelle des monochromatischen Lichtes Lichtwellen mit mehreren Frequenzen einstrahlt. Aus der Gleichung für P_{NL} ist ersichtlich, daß dann verschiedene Anteile bei den *Summen-* und *Differenzfrequenzen* entstehen und demgemäß von

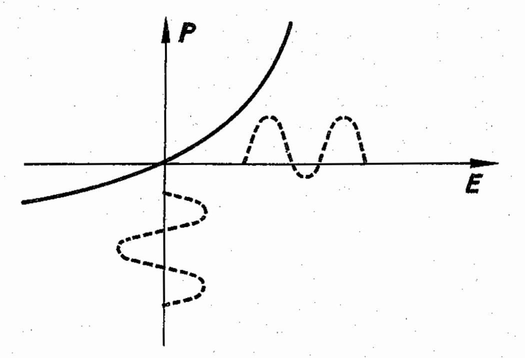


Abb. 4.11.2–1 Nichtlineare Abhängigkeit der Polarisation P von der Laserfeldstärke E

dem Material Felder entsprechender Frequenz abgestrahlt werden können.
In vielen Materialien treten aus Symmetriegründen bestimmte Beiträge in der nichtlinearen Entwicklung der Polarisation nach der Feldstärke nicht auf, d. h. die entsprechenden *nichtlinearen Suszeptibilitäten* $\varkappa_i$ verschwinden. So gilt in allen *optisch isotropen Stoffen* $\varkappa_2 = 0$, was bedeutet, daß nichtlineare Effekte zweiter Ordnung nur in bestimmten Kristallen auftreten können.

4.11.2.2. Nichtlineare Effekte und ihre Anwendung. Die Erzeugung der zweiten Harmonischen (vgl. 4.11.2.1.) wurde als erster nichtlinearer optischer Effekt 1961 von P. A. FRANKEN beobachtet. Bei diesem Experiment fiel das Licht eines Rubinlasers ($\lambda = 0{,}694\,\mu$m) auf einen Quarzkristall. Hinter dem Kristall konnte Licht der doppelten Frequenz ($\lambda = 0{,}347\,\mu$m) nachgewiesen werden (Abb. 4.11.2–2).

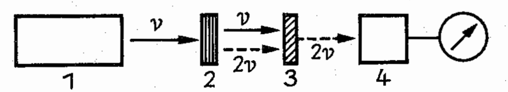


Abb. 4.11.2–2 Schema der Anordnung zur Erzeugung der zweiten Harmonischen in Quarz; 1 Rubinlaser, 2 Quarzkristall, 3 Filter, das die Grundwelle sperrt, 4 fotoelektrischer Empfänger mit nachfolgender elektronischer Nachweiseinrichtung

Für die Intensität der zweiten Harmonischen bei Einstrahlung einer Grundwelle der Intensität $I(\nu)$ auf eine Kristallplatte der Dicke l folgt aus dem Ausdruck für die nichtlineare Polarisation und den Maxwell-Gleichungen

$$I(2\nu) \sim \varkappa_2 I^2(\nu) \sin^2(\pi \Delta q l)/(\pi \Delta q)^2$$

mit Wellenzahl $q(\nu) = n(\nu)\,\nu/c$ und $\Delta q = q(2\nu) - 2\,q(\nu)$

($n(\nu)$ = Brechzahl). Man sieht, daß die Intensität bei der Frequenz 2ν quadratisch mit der Intensität der Grundwelle anwächst; diese Abhängigkeit gilt allerdings nur für kleine Umwandlungsraten, bei hohen Umwandlungsraten ist das Wachstum infolge des Abbaues der Grundwelle schwächer ausgeprägt. Für dünne Kristalle steigt $I(2\nu)$ mit der Kristallänge ebenfalls quadratisch an. Insgesamt wächst $I(2\nu)$ bis zur Länge $l_K = 1/(2\Delta q)$, anschließend wird die Energie bis zur Länge $2\,l_K$ wieder in die Grundwelle zurückgepumpt. Die Größe l_K wird als *Phasenkohärenzlänge* bezeichnet; auf dieser Länge ändert sich infolge der *Dispersion* des Materiales die relative Phase zwischen der Feldstärke bei 2ν und der durch die Grundwelle bei ν erzeugten Polarisation so, daß der Gewinn in einen Verlust umschlägt (Abb. 4.11.2–3). Eine besonders effektive Erzeugung der Harmonischen ist im Falle $\Delta q = 0$ möglich, den man als *Phasenanpassung* bezeichnet, weil die günstige Phasenlage zwischen Polarisation und Feldstärke für beliebige Kristalllängen erhalten bleibt.
Experimentell erreicht man eine *Phasenanpassung in doppelbrechenden Kristallen* (vgl. 2.5.2.3.). Hier können sich bei einer Frequenz zwei Wellen mit unterschiedlicher Phasengeschwindigkeit *c/n* als *ordentliche* (o) und *außerordentliche Welle* (ao) ausbreiten, wobei die Phasengeschwindigkeit bzw. Brechzahl der außerordentlichen Welle richtungsabhängig ist und nur in Richtung der optischen Achse mit den Werten für die ordentliche Welle übereinstimmt. Beide Wellen sind senkrecht zueinander polarisiert. Bei geeigneter Wahl des Kristalles und der Durchstrahlungsrichtung kann beispielsweise $n_{ao}(2\nu) = n_o(\nu)$ und damit $\Delta q = 0$ erreicht werden. Durch hohe Nichtlinearitäten und Möglichkeiten zur Phasenanpassung zeichnen sich z. B. die Kristalle Lithiumniobat, Lithiumiodat und Kaliumdikydugenphosphat aus. Es konnten bereits Umwandlungsraten von nahezu 100% realisiert werden.
Höhere Harmonische können infolge der nichtlinearen Glieder höherer Ordnung in der Polarisation ($\varkappa_3 E^3$, $\varkappa_4 E^4$ usw.) in analoger Weise erzeugt werden. Für die Erzeugung kurzwelliger Strahlung ist es jedoch zumeist günstiger, mehrmals nacheinander die zweite Harmonische zu bilden.

Frequenzmischung. Bei Einstrahlung von zwei monochromatischen Wellen der Frequenz ν_1 bzw. ν_2 entstehen in einem nichtlinear optischen Kristall nichtlineare Polarisationsanteile bei den Frequenzen $2\nu_1$ und $2\nu_2$, die die Erzeugung der zweiten Harmonischen (vgl. 4.11.2.2.) bewirken, sowie bei den Frequenzen $\nu_3 = \nu_1 + \nu_2$ und $\nu_4 = |\nu_1 - \nu_2|$, die den Aufbau von elektromagnetischen Wellen bei dieser *Summen-* bzw. *Differenzfrequenz* verursachen. Eine sehr effektive Erzeugung dieser Wellen wird analog zur Erzeugung der Harmonischen im Falle der Phasenanpassung erreicht, die bei kolinearer Einstrahlung die Erfüllung von $\Delta q = q_{3,4} - q_1 - q_2 = [\nu_{3,4} n(\nu_{3,4}) - \nu_1 n(\nu_1) - \nu_2 n(\nu_2)]/c = 0$ erfordert.
Durch Frequenzmischung kann man kohärente Strahlung in anderen Spektralbereichen erzeugen, wobei besonders die Generation sehr langwelliger Strahlung im *fernen Infrarot* eine wichtige Rolle spielt. Zum anderen kann man die Wellenlängentransformation – z. B. vom IR- in den sichtbaren Bereich – ausnutzen, um *optische Signale* in Spektralbereichen nachzuweisen, in denen besonders empfindliche oder besonders schnell ansprechende Fotoempfänger zur Verfügung stehen.

Parametrische Verstärkung bzw. Generation ist das Anwachsen der Intensität bzw. die Erzeugung zweier Lichtwellen der Frequenzen ν_1 und ν_2 in einem nichtlinear optischen Kristall bei Einstrahlung einer starken Pumpwelle bei der Frequenz $\nu_3 = \nu_1 + \nu_2$. Es handelt sich also um eine Umkehrung der *Frequenzmischung*, wobei dementsprechend die Phasenanpassung $\Delta q = q(\nu_3) - q(\nu_1) - q(\nu_2) = 0$ erforderlich ist. Man kann den Prozeß auch in der Weise interpretieren, daß die Wellen bei den Frequenzen ν_1 und ν_2 in einem Material in Wechselwirkung stehen, indem ein Parameter, hier die Brechzahl, durch Einwirkung der starken Pumpwelle periodisch moduliert wird; dies wird in Analogie zur Hochfrequenzphysik *parametrischer Prozeß* genannt.
Parametrische Generatoren werden erfolgreich als kohärente Lichtquellen mit abstimmbarer Frequenz im IR- und sichtbaren Spektralbereich eingesetzt. Die *Frequenzabstimmung* wird durch eine Veränderung der Phasenanpassung erreicht, was z. B. durch Änderung von Temperatur oder Orientierung der nichtlinear optischen Kristalle möglich ist.

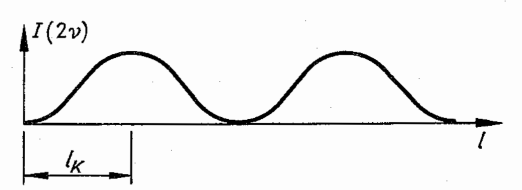


Abb. 4.11.2–3 Intensität der zweiten Harmonischen $I\,(2\nu)$ in Abhängigkeit von der Kristallänge l

Selbstfokussierung und optischer Kerr-Effekt. Aus dem nichtlinearen Polarisationsanteil dritter Ordnung (vgl. 4.11.2.1) folgt eine Intensitätsabhängigkeit der Brechzahl. Wächst sie mit der Intensität, so werden die Strahlen eines Lichtbündels in den Bereich hoher Intensitäten im Bündelzentrum gelenkt, wodurch sich das Bündel einschnürt. Man bezeichnet diesen Effekt als *Selbstfokussierung.*
Man kann diese Brechzahländerung auch mit einer anderen Lichtquelle untersuchen, wobei i. allg. eine Abhängigkeit von der Polarisationsrichtung, also eine *Doppelbrechung,* festgestellt wird, die man als *optischen* oder *laserinduzierten* Kerr-Effekt bezeichnet. Auf der Grundlage dieses Effektes können bei Anwendung von ultrakurzen Pumplichtimpulsen *optische Schalter* mit Schaltzeiten im Pikosekundenbereich aufgebaut werden.
Nichtlineare Resonanzeffekte. Wenn die *Summen-* oder *Differenzfrequenzen* der eingestrahlten Lichtwellen mit *Übergangsfrequenzen* $\nu_{ij} = (E_i - E_j)/h$ der atomaren Systeme übereinstimmen, so können diese Systeme durch eine *Mehrphotonenwechselwirkung* angeregt werden. Beispiele für solche resonanten Wechselwirkungsprozesse sind die *Mehrphotonenabsorption* und *-emission* (speziell die Zweiphotonenabsorption und -emission) und die stimulierte Raman-Streuung (Abb. 4.11.2–4). Bei der *Zweiphotonenabsorption* wird je ein Photon der Frequenzen ν_1 und ν_2 vernichtet und dabei ein Atom vom Energieniveau 0 zum Niveau 1 angeregt, wobei $h\nu_1 + h\nu_2 = E_1 - E_0$ gilt. Die *Zweiphotonenemission* stellt den Umkehrungsprozeß dar, bei dem das Atom in den tieferen Zustand übergeht und gleichzeitig zwei Photonen emittiert werden. Beim *spontanen* und *stimulierten Raman-Effekt* werden bei Wechselwirkung mit atomaren Systemen im Grundzustand Photonen der größeren Frequenz ν_1 vernichtet und Photonen der kleineren Frequenz ν_2, die als *Stokes-Frequenz* bezeichnet wird, erzeugt, wobei gleichzeitig das atomare System, z.B. ein Molekül, angeregt wird. Es gilt der Energieerhaltungssatz in der Form $h(\nu_1 - \nu_2) = E_1 - E_0$. Beim *stimulierten Raman-Effekt* treten in der Probe zwei starke, kohärente Lichtwellen der Frequenz ν_1

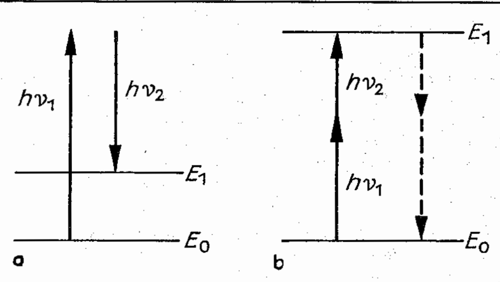


Abb. 4.11.2–4 Übergänge bei (a) der stimulierten Raman-Streuung und (b) der Zweiphotonenabsorption bzw. -emission

und ν_2 in Wechselwirkung, wodurch eine kohärente Anregungswelle der Frequenz $\nu_1 - \nu_2$ entsteht. In Festkörpern sind diese Anregungswellen im Kristall sich wellenförmig ausbreitende optische Schwingungen der Gitterbausteine. Beim stimulierten Raman-Effekt werden *optische Phononen* der Frequenz $\nu_{\text{Phonon}} = \nu_1 - \nu_2$ erzeugt. Die Wechselwirkung dieser Anregungswelle mit den Lichtwellen führt zu einer phasentreuen Verstärkung der niederfrequenten Welle. Bei niedrigen Lichtfeldstärken kann keine kohärente Anregungswelle erzeugt werden; dann tritt lediglich eine schwache Streuung an den *inkohärenten Nullpunktschwingungen* der atomaren Systeme auf, die als *spontaner Raman-Effekt* bezeichnet wird.
Bei starker Einstrahlung kann selbst bei einer Einphotonenabsorption (vgl. 4.11.1.1.) eine nichtlineare Abhängigkeit zwischen Polarisation und Lichtfeldstärke entstehen. Die Nichtlinearität wird in diesem Falle durch eine Intensitätsabhängigkeit der Besetzungszahlen der atomaren Energieniveaus bewirkt. Bei hohen Lichtintensitäten können sich relativ viele Teilchen im angeregten Zustand befinden. Dadurch wird die Absorption der Probe verkleinert. Diesen Effekt bezeichnet man als *Absorptionssättigung.*

5. Physik der Elementarteilchen und Felder, Kerne, Atome und Moleküle

5.1. Elementarteilchen

Die Physik der Elementarteilchen ist im wesentlichen eine Physik hoher Energien. Die überwiegende Anzahl der Teilchen tritt unter natürlichen irdischen Bedingungen nicht auf. Um diese Teilchen überhaupt im Laboratorium zu erzeugen, müssen das ursprünglich vorliegende bzw. relativ leicht zu gewinnende Elektron und Proton auf Energien im Gigaelektronenvoltbereich beschleunigt und auf ein materielles Target oder einen zweiten Teilchenstrahl geschossen werden. Die Erforschung der zwischen Elementarteilchen ablaufenden Streu- und Umwandlungsprozesse und die Suche nach neuen Teilchen mit größerer Ruhmasse erfordert den Bau immer leistungsfähigerer Teilchenbeschleuniger. In den letzten Jahren sind allerdings auch äußerst wichtige Experimente angelaufen, die unabhängig von Beschleunigern durchgeführt werden. Ein herausragendes Beispiel stellen die Anlagen zum Nachweis des Protonzerfalls (vgl. 5.1.4.3.) dar.

5.1.1. Teilchen und Felder

Unter dem Begriff Elementarteilchen faßt man allgemein die Gesamtheit der subatomaren Teilchen zusammen, deren bekannteste Vertreter Photon, Elektron, Proton und Neutron sind. Solange nur diese wenigen Teilchen bekannt waren, durch deren Zusammenwirken Struktur und Verhalten der Atome und Atomkerne mit Ausnahme des β-Zerfalls vollständig verstanden werden können, schien die Situation einfach. Die Vielfalt der Elementarteilchen deutete sich zuerst in der kosmischen Höhenstrahlung an, wo das Positron (1932), das Myon und das Pion (1947) beobachtet wurden. Eine systematische Erforschung dieser Teilchen begann mit dem Einsatz von Beschleunigern; dabei wurde auch eine Vielzahl bis dahin nicht bekannter Teilchen gefunden. Nachdem diese Entwicklung schon zu einem Abschluß gekommen zu sein schien, leitete die Entdeckung des Tau-Leptons und der neuen Charm-Teilchen (1974) eine Bewegung ein, die erst kürzlich (1984) mit der Etablierung von drei Familien fundamentaler Teilchen einen Höhepunkt erreicht hat. Die Elementarteilchenphysik unterscheidet sich hinsichtlich der umgesetzten Energien deutlich von anderen Gebieten der Mikrophysik: sind es bei Atomen und Molekülen Elektronenvolt und bei Atomkernen Megaelektronenvolt, so handelt es sich hier um Gigaelektronenvolt, ja neuerdings um Hunderte von Gigaelektronenvolt. Einem großen Teilchenimpuls p entspricht, im Sinne des Welle-Teilchen-Dualismus nach der de-Broglie-Beziehung, die sehr kleine Wellenlänge $\lambda = h/p$, die wie in der Optik eine hohe Auflösung ermöglicht. Der Einsatz höherer Energien erlaubt demnach das Erfassen immer kleinerer Strukturen. Mit dem Begriff der Elementarteilchen verbindet sich also keineswegs die Vorstellung kleinster, nicht weiter auflösbarer Objekte.
Die meisten Teilchen sind instabil, sie zerfallen und wandeln sich dabei in andere Teilchen um. Auch bei der Wechselwirkung von Teilchen miteinander enstehen i. allg. völlig neue Teilchen, darunter bei ausreichend hohen kinetischen Energien der Stoßpartner auch schwerere Teilchen als ursprünglich vorhanden waren. Diese Prozesse können i. allg. nicht als bloße Umverteilung von Grundbausteinen aufgefaßt werden wie etwa die Kernreaktionen. Die Elementarprozesse der Teilchenphysik sind die der Erzeugung und Vernichtung von Teilchen. Die Gesetze der Teilchenphysik werden durch die Untersuchung von Streu-, Erzeugungs- und Zerfallsprozessen studiert. Die Ergebnisse dieser Experimente werden quantitativ durch die Angabe von Wirkungsquerschnitten, Lebensdauern und Verzweigungsverhältnissen ausgedrückt. Die Untersuchungen zeigen, daß nicht alle mit den bekannten Erhaltungssätzen für Energie, Impuls und elektrische Ladung verträglichen Prozesse beobachtet werden. Es ist möglich, den Teilchen gewisse diskrete Quantenzahlen als für sie charakteristische Größen so zuzuordnen, daß die möglichen Prozesse durch Erhaltungssätze für diese neuen Größen von den nicht möglichen unterschieden werden. Außerdem zeigt die quantitative Auswertung, daß die Wechselwirkungsprozesse zwischen Elementarteilchen von sehr verschiedener „Intensität" sind, charakterisiert durch die Größe der Wirkungsquerschnitte bzw. die Kleinheit der Lebensdauer instabiler Teilchen. Diese Erfahrungen haben zur Klassifikation der Teilchen in Hadronen und Leptonen geführt und zur Unterteilung der Wechselwirkungen in starke, elektromagnetische und schwache Wechselwirkungen. Die Elementarteilchen unterliegen nach heutiger Kenntnis voll und ganz den Gesetzen von Quanten- und Relativitätstheorie. Daraus ergibt sich eine wesentliche Einschränkung für die Anwendbarkeit anschaulicher Vorstellungen, die der Makrowelt angehören. Den Elementarteilchen dürfen daher nur solche Eigenschaften zugeordnet werden, für deren Existenz ein direkter experimenteller Beleg vorliegt, nicht solche, die aus hypothetischen Strukturvorstellungen folgen. Nur bei einigen wenigen Teilchen kann man begründet von einer räumlichen Ausdehnung sprechen. Im Einklang mit den Prinzipien der Quantenphysik sind die Teilchen ein und derselben Sorte ununterscheidbar. Darüber hinaus fordert die relativistische Quantenphysik, daß zu jeder Sorte ein *Antiteilchen* auftritt. Die Umwandlungsprozesse zwischen Elementarteilchen tragen statistischen Charakter.
Die Physik der Elementarteilchen ist und bleibt ein prinzipiell nicht abgeschlossenes Gebiet. Als gesicherte Tatsachen können zuerst nur die durch wiederholte, unabhängige Experimente aufgefundenen Eigenschaften der schon bekannten Teilchen angesehen werden, dazu auch in gewissem Maße einige allgemeine theoretische Gesetzmäßigkeiten. Neue Einsichten können sich aus Experimenten ergeben, die entweder durch den Einsatz höherer Beschleunigungsenergien die Auflösung feinerer Strukturen gestatten, oder aber durch eine wesentliche Steigerung der Genauigkeit der Messungen gängige und durchaus plausible Vorstellungen korrigieren. Beispiele hierfür bieten die Experimente zur Bestimmung der Elektron-Neutrinomasse und zum Nachweis des Protonzerfalles.

5.1.1.1. Nachweis von Elementarteilchen. Nachweismethoden. Nur die wenigsten Teilchen können unmittelbar in den Nachweisgeräten beobachtet werden. Meist lassen sich nur die Zerfallsprodukte nachweisen und quantitativ erfassen, um aus ihnen dann einen Rückschluß auf das ursprüngliche Teilchen zu ziehen. Es können entweder die elektrisch geladenen, langlebigen oder stabilen Teilchen p, $e^{\pm}$, $\mu^{\pm}$, $\pi^{\pm}$, $K^{\pm}$ oder Photonen nachgewiesen werden. Bei geladenen Teilchen wird der Umstand ausgenutzt, daß sie beim Durchgang durch Materie entweder eine Spur angeregter oder ionisierter Zentren (Atome oder Moleküle) längs ihrer Flugbahn hinterlassen oder zur Ausstrahlung von Licht führen. Im Falle von Photonen macht man sich den Effekt der Konversion zunutze: Hochenergetische Photonen gehen im elektrischen Feld schwerer Atome in Elektron-Positron-Paare über, die dann wie andere geladene Teilchen analysiert werden. Das Ziel der Messung ist sowohl die Identifikation des Teilchens als auch die Messung seines Impulses. Zu diesem Zweck setzt man die verschiedenen Detektoren in ein konstantes Magnetfeld und bestimmt aus der Bahnkrümmung den Impuls der Teilchen. Die Indentifikation des Teilchens läuft auf die Ermittlung seiner Masse hinaus. Dazu muß zusätzlich zum Impuls entweder die Energie oder die Geschwindigkeit bestimmt werden. Die Geschwindigkeitsbestimmung erfolgt z. B. durch Ausnutzung des Čerenkov-Effektes.

Nachweisgeräte. Die *Blasenkammer* ist ein Gefäß, das eine überhitzte Flüssigkeit, z. B. Wasserstoff, Deuterium oder Propan, enthält. Das hindurchfliegende Teilchen hinterläßt auf seiner Flugbahn ionisierte Atome, die den Siedevorgang einleiten. Die entsprechenden Gasbläschen kennzeichnen die Spur des Teilchens, die fotografiert wird, um sie später zu analysieren. Die Blasenkammer hat die noch früher entwickelte *Nebelkammer* ersetzt, deren Wirkungsweise auf dem Prinzip des übersättigten Dampfes beruht.

Bei *Streamerkammern* werden die sekundären Ionisationseffekte bei Vorhandensein eines kurzzeitig, d.h. $\approx 10^{-9}$ s lang, angelegten, sehr starken elektrischen Feldes ausgenutzt, durch das die primär erzeugten Elektronen und Ionen beschleunigt werden. Es kommt längs der Teilchenbahn zu bandförmigen Leuchterscheinungen, die entweder direkt oder unter Zuhilfenahme von Lichtverstärkern fotografiert werden.

Eine andere Klasse von Detektoren, die *Ionisations-, Proportional-* und *Driftkammern*, kommt ohne eine fotografische Aufzeichnung der Teilchenspur aus. Auch hier erzeugt das nachzuweisende Teilchen im Gas der Kammer Elektronen und Ionen. Ein elektrisches Feld sorgt für Trennung und Beschleunigung dieser Ladungsträger, so daß durch sekundäre Ionisation eventuell weitere erzeugt werden. Die entstandenen Elektronen werden durch ein System paralleler Drähte aufgefangen. Diese Drähte sind so in Ebenen (z.B. parallel zur x-y-Ebene eines rechtwinkligen Koordinatensystems) gespannt, daß immer abwechselnd die Drähte einer Ebene längs der x-Achse und die der folgenden Ebene längs der y-Achse verlaufen. Eine Registrierung des elektrischen Impulses erfolgt durch denjenigen Draht, welcher der Teilchenspur am nächsten ist. Ein einzelnes Paar solcher dicht hintereinander aufgestellter Drahtebenen fixiert also den Schnittpunkt der Spur, weitere Paare ermöglichen die Bestimmung des Kurvenverlaufes. Die gesamte, in Form von elektrischen Impulsen anfallende Information wird unmittelbar in einem Computer verarbeitet.

Die Bestimmung der Geschwindigkeit ist mit Hilfe von *Čerenkov-Zählern* möglich, deren Wirkungsweise auf dem Čerenkov-Effekt beruht: Elektrisch geladene Teilchen strahlen bei Durchgang durch ein Medium mit dem Brechungsindex n Licht aus, falls die Bedingung $v/c > 1/n$ erfüllt ist. Durch eine Staffelung von Medien mit unterschiedlichem Brechungsindex oder eine Messung des Abstrahlungswinkels der Čerenkov-Strahlung läßt sich ein Rückschluß auf die Geschwindigkeit des hindurchfliegenden Teilchens ziehen.

Neben diesen Geräten, die der Messung eines einzelnen Teilchens dienen, gibt es sogenannte *Kalorimeter* zur Bestimmung der Energie, die ein manchmal aus vielen Teilchen bestehendes Teilchensystem hat. Die primären Teilchen übertragen ihre Energie auf einen Schauer geladener Teilchen, die ihrerseits ein Szintillatormaterial zur Lichtaussendung anregen. Die Lichtausbeute ist ein Maß für die Energie des ursprünglichen Teilchensystems. Auf diese Weise läßt sich die Energie eines Neutrinos bestimmen, das in einer Reaktion erzeugt worden ist. Sie ergibt sich als Differenz aus der Anfangsenergie der Reaktionspartner (etwa Myon und Proton) und der im Kalorimeter gemessenen Energie aller gleichzeitig mit dem Neutrino erzeugten Teilchen. Bei den neueren Experimenten der Elementarteilchenphysik werden sehr komplexe Apparaturen eingesetzt, die aus einer Kombination vieler einzelner Detektoren bestehen.

5.1.1.2. Meßbare Eigenschaften. Masse. Genau wie ein makroskopischer Körper haben die Teilchen eine Masse. Die für das Teilchen charakteristische Ruhmasse wird gewöhnlich in der Einheit GeV/c^2 angegeben, wobei die Einsteinsche Beziehung $E = mc^2$ ausgenutzt wird. Das Massenspektrum der Elementarteilchen kann mit den gegenwärtig existierenden Beschleunigern bis in den Bereich von 100 GeV/c^2 untersucht werden (1 GeV/c^2 = $1{,}782 \cdot 10^{-27}$ kg).

Spin. Der Eigendrehimpuls eines Teilchens heißt Spin. Er wird in Einheiten der Planckschen Konstante $\hbar = h/2\pi$ gezählt. Im Einklang mit der Quantentheorie nimmt der Spin J nur halb- oder ganzzahlige Werte an. Masse und Spin charakterisieren immer eine Teilchensorte, etwa das Proton. Ein konkret existierendes Proton wird erst durch die zusätzliche Angabe von Impuls p und Spinkomponente J_z in Richtung einer willkürlich gewählten z-Achse vollständig bestimmt. Dabei kann $J_z/\hbar$ die $2J+1$ Werte $-J$, $-J+1$, ..., J annehmen. Diese Regel gilt nicht für masselose Teilchen. So spielen die Photonen, für die $J = 1$ gilt, eine Sonderrolle. Nur die Spinprojektion in Richtung des Impulses ist meßbar, sie nimmt nur die zwei Werte $+1$ oder -1 an. Der Spin kann für langlebige Teilchen aus einem Stern-Gerlach-Experiment bestimmt werden; i. allg. erschließt man ihn aus einer Analyse der Winkelabhängigkeit der Zerfallsprozesse des betreffenden Teilchens.

Lebensdauer. Für relativ langlebige Teilchen, die in den Nachweisapparaturen noch eine beobachtbare Spur hinterlassen, kann die Lebensdauer τ direkt aus der Länge der Spur ermittelt werden. Für sehr kurzlebige Teilchen, die sogenannten Resonanzen, ist das unmöglich. Da sich die Resonanzen als Maxima im Wirkungsquerschnitt bemerkbar machen (Abb. 5.1.1–1), kann aus der energetischen Breite Γ des Maximums gemäß der quantentheoretischen Beziehung $\Gamma\tau \approx \hbar \approx 6{,}6 \cdot 10^{-22}$ MeV · s auf die mittlere Lebensdauer geschlossen werden. Für die Resonanz Δ (1236), die eine Breite von 120 MeV hat, findet man auf diese Weise $\tau \approx 5{,}5 \cdot 10^{-24}$ s.

Elektrische Ladung. Die elektrische Ladung aller bisher gefundenen Teilchen ist ein positives oder negatives Vielfaches der Elementarladung e. In dieser Einheit hat das Elektron die Ladung $Q = -1$, das Proton $Q = +1$ und das Neu-

tron $Q = 0$. Im Falle von Teilchen, die unterschiedlich geladen auftreten können, wird die Ladung rechts oben am Symbol angegeben, z.B. Π^+, Π^0. Die hypothetischen Quarkteilchen sollten die Ladungen $Q = \pm 2/3$ oder $Q = \pm 1/3$ tragen.

Magnetisches Moment. Teilchen mit Spin haben i.allg. ein magnetisches Moment, d.h. sie verhalten sich in einem Magnetfeld wie eine kleine Magnetnadel. Sofern diese Teilchen frei fliegend zu beobachten sind, läßt sich ihr magnetisches Moment durch eine geeignete Anordnung von Magnetfeldern experimentell bestimmen. Das magnetische Moment μ ist die Summe aus einem sogenannten normalen Anteil $e\hbar/2M$, das jedes elektrisch geladene Teilchen der Masse M nach der relativistischen Quantenmechanik besitzt, und einem anormalen Anteil, der Resultat der komplizierten Wechselwirkungen der Teilchen untereinander ist. Präzisionsmessungen von μ ergaben für Elektronen eine Abweichung von 1,1596 ‰, für Myonen von 1,1659 ‰ vom normalen Wert. Das befindet sich im Einklang mit den in der Quantenelektrodynamik errechneten Korrekturen. Die Momente von Proton, Neutron u.a. Baryonen weichen viel stärker von den normalen Werten ab: $\mu_p = 2{,}793\, e\hbar/2M_p$, $\mu_n = -1{,}913\, e\hbar/2M_p$. Darüber hinaus ist μ bisher nur für die Baryonen Λ, Σ^+, Σ^-, Ξ^0 und Ξ^- gemessen worden.

Ausdehnung von Teilchen kann durch die räumliche Verteilung ihrer elektrischen Ladung definiert werden. Die experimentelle Bestimmung dieser Ladungsdichte erfolgt durch elastische Streuung von Elektronen. Für Proton und Neutron ergeben sich Radien in der Größenordnung von $1\,\text{fm} = 10^{-13}\,\text{cm}$.

5.1.1.3. Ladungsartige Quantenzahlen. Die Charakterisierung von Teilchen erfordert über die bisher angeführten Eigenschaften hinaus die Angabe gewisser Quantenzahlen, die keine Rolle in der Makrophysik spielen und daher unanschaulich sind. Sie sind für das Einzelteilchen nicht direkt meßbar, sondern dienen zur Klassifizierung der Teilchen und zur Formulierung von Einschränkungen bei Umwandlungsprozessen.

Baryonzahl. Den Baryonen (vgl. 5.1.2.2.) wird die Baryonzahl $B = 1$ zugeordnet, ihren Antiteilchen $B = -1$ und allen übrigen Teilchen $B = 0$. Demnach hat das Proton $B = +1$, Elektron, Photon und Pion haben $B = 0$. B wurde als eine Quantenzahl eingeführt, für welche ein absoluter Erhaltungssatz (vgl. 5.1.3.1.) angenommen wurde. Dadurch

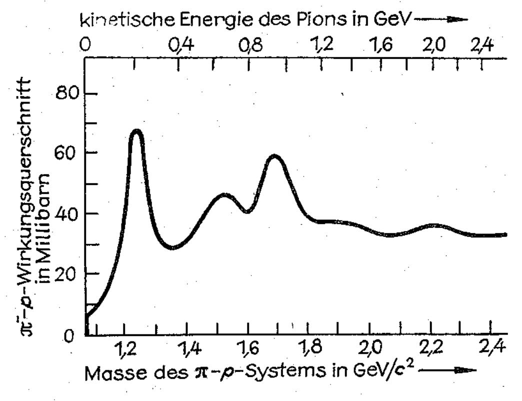


Abb. 5.1.1-1 Totaler Wirkungsquerschnitt der Streuung von negativen Pionen an Protonen

werden u.a. Reaktionen von der Art $p + p \to \pi^+ + \pi^+$ oder $p + e^- \to 2\gamma$ ausgeschlossen und damit die Stabilität der massiven Materie theoretisch erfaßt. Allerdings geben die Vorstellungen von einer möglichen Vereinigung der schwachen, elektromagnetischen und starken Wechselwirkungen (vgl. 5.1.3.7.) Veranlassung, diesen Erhaltungssatz einer experimentellen Prüfung zu unterziehen. Es wird insbesondere nach Zerfällen des Protons gesucht, um eine eventuelle schwache Verletzung der Erhaltung von B quantitativ zu bestimmen.

Leptonzahlen. Jedem Paar von Leptonen (vgl. 5.1.2.1.) ist eine eigene Leptonzahl zugeordnet. So haben das Elektron e^- und sein Neutrino ν_e die Elektronenzahl $L_e = +1$, ihre Antiteilchen e^+ und $\bar{\nu}_e$ aber $L_e = -1$. Entsprechend ordnet man den anderen Leptonen μ, τ Leptonzahlen zu. Der Erhaltungssatz für L_e oder L_μ verbietet beispielsweise den Zerfall $\mu^- \to e^- + \gamma$ oder den Prozeß $\bar{\nu}_e + n \to p + e^-$. Auch der Erhaltungssatz für die Leptonzahlen hat eher den Charakter einer Arbeitshypothese, die immer von Neuem einer experimentellen Prüfung unterzogen werden muß.

Hyperladung. Die Hyperladung Y dient der weiteren Unterscheidung von Hadronen (vgl. 5.1.2.2.). Die möglichen Werte von Y sind ganzzahlig. Die Einführung der Hyperladung hat sich als zweckmäßig erwiesen, um die Gesetzmäßigkeiten bei der Erzeugung der K-Mesonen und der Baryonen Σ, Λ, Ξ zu erfassen.

Charm. Die Entdeckung der neuen Teilchen ψ, D, ... hat die Einführung der Quantenzahl Charm (C) nahegelegt, die gleichsfalls nur für Hadronen definiert ist. Allen vorher bekannten Hadronen wird der Wert $C = 0$ zugeordnet. Im Rahmen des Quarkmodells bedeutet dies, ein viertes Quarkteilchen mit $C = +1$ einzuführen, während gleichzeitig die ursprünglichen Quarkteilchen die Quantenzahl $C = 0$ erhalten. Charm und Hyperladung spielen eine weitgehend ähnliche Rolle als Ladungen, die nur in der Mikrowelt von Bedeutung sind. Die Entdeckung der schweren Mesonen Y, B, ... usw. sowie die erst 1984 gefundenen Anzeichen einer weiteren Teilchenart (vgl. 5.1.2.3.) machen die Einführung weiterer Quantenzahlen dieser Art erforderlich.

Strangeness. Alternativ zur Hyperladung benutzt man auch die Quantenzahl Strangeness S [engl., Seltsamkeit], die mit Y und B durch die Beziehung $S = Y - B$ verknüpft ist. Die Bedeutung von S ergibt sich daraus, daß eines der Quarkteilchen (s) die Quantenzahl $S = -1$, die übrigen aber $S = 0$ tragen. Zur Kennzeichnung der neuen Hadronen verwendet man die Quantenzahlen S, C, usw.

5.1.1.4. Isospin. Im Massenspektrum der Hadronen zeigen sich Multipletts von Teilchen mit relativ kleinen Massenunterschieden, die sich durch ihre elektrische Ladung voneinander unterscheiden, aber den gleichen Spin und übereinstimmende Quantenzahlen B, Y, C aufweisen. Beispiele dafür sind die Dubletts (p, n) und (K^+, K^0), die Tripletts (Σ^+, Σ^0, Σ^-) (Π^+, Π^0, Π^-) und das Singulett Λ, das keine Multiplettpartner hat. Es liegt nahe und ist am Beispiel des Nukleondubletts zuerst durchgeführt worden, die zu einem Multiplett gehörenden Teilchen als verschieden geladene Varianten eines Teilchens aufzufassen. Der mathematischen Durchführung dieser Konzeption dient der Formalismus des Isospins, der analog zur Quantentheorie des Drehimpulses aufgebaut wird: Ein in einem abstrakten Raum definierter Isospinvektor mit einem Quadrat $I(I+1)$ hat $2I+1$ verschiedene Einstellmöglichkeiten, wobei eine Komponente $I_3 = -I, -I+1, \ldots +I$ insgesamt $2I+1$ verschiedene Werte durchläuft. Benachbarten Werten von I_3

entsprechen Teilchen, deren elektrische Ladungen Q sich um 1 unterscheiden. Q ist durch I_3, Baryonzahl B, Strangeness S und Charmquantenzahl C bestimmt: $Q = I_3 + (B + S + C)/2$. Ebenso wie der Spin J nimmt auch der Isospin I nur halb- oder ganzzahlige Werte an; alle Teilchen eines Multipletts haben gleichen Isospin. $I = 0$ liefert ein Singulett, $I = 1/2$ ein Dublett, $I = 1$ ein Triplett usw.

5.1.1.5. Parität. Wenn einmal von den Neutrinos abgesehen wird, so ist allen Teilchen eine Parität P zugeordnet, die die Werte $+1$ oder -1 annimmt. Bei einer Raumspiegelung geht z. B. die quantentheoretische Wellenfunktion X eines Zweiteilchensystems mit dem Bahndrehimpuls l bis auf ein Vorzeichen in sich über: $X \to (-1)^l P_1 P_2 X$. Auf diese Weise gehen die Paritäten P_1, P_2 der beteiligten Teilchen in das Transformationsgesetz ein. Die Paritäten lassen sich aus der Winkelabhängigkeit der Wirkungsquerschnitte berechnen.

5.1.1.6. Antiteilchen. Zu jedem Teilchen gibt es ein Antiteilchen, dessen Quantenzahlen in einer typischen Beziehung zu denen des Teilchens stehen (Tab. 5.1.1–2). Das Antiteilchen wird durch einen Querstrich über dem Symbol bezeichnet, z. B. $\bar{p}$. Abweichend von dieser Regel bezeichnet e^+ das Antielektron oder Positron, μ^+ das Antimyon. Falls alle landungsartigen Quantenzahlen eines Teilchens den Wert Null haben, ist dieses mit seinem Antiteilchen identisch, z. B. $\bar{\Pi}^0 = \Pi^0$, $\bar{\gamma} = \gamma$. Die Antiteilchen stimmen in Masse und Lebensdauer mit ihrem Partner überein. Insbesondere sind $\bar{p}$ und e^+ stabile Teilchen. Ihre scheinbare Instabilität im Laboratorium rührt daher, daß sie in einer Umgebung, die Protonen und Elektronen enthält, mit großer Wahrscheinlichkeit einer Paarvernichtung unterliegen.

Tab. 5.1.1–2 Beziehungen zwischen Teilchen und zugehörigen Antiteilchen

Teilchen	Antiteilchen
B	$-B$
L_e	$-L_e$
L_μ	$-L_\mu$
Q	$-Q$
I_3	$-I_3$
Y	$-Y$
C	$-C$
P	$(-1)^{2J}P$
J	J
I	I

Es ist von vornherein nicht auszuschließen, daß in anderen Teilen des Universums Antimaterie in makroskopischem Ausmaß existiert, die aus Antikernen, Antiatomen, Antimolekülen usw. besteht, wofür allerdings astrophysikalische Hinweise fehlen. Eine für die Teilchenphysik charakteristische Besonderheit besteht darin, daß zu jedem Prozeß ein anderer möglicher Prozeß gehört, in dem ein oder mehrere Teilchen aus dem Anfangs- in den Endzustand oder umgekehrt gewechselt haben und nun als Antiteilchen erscheinen; z. B. ist mit $a + b \to c + d$ auch $a + \bar{c} \to \bar{b} + d$ ein möglicher Prozeß.

Die Tab. 5.1.1–2 zeigt, daß jedes aus Teilchen und Antiteilchen bestehende System die Quantenzahlen $B = L_e = L_\mu = L_\tau = 0$, $Q = I = Y = C = 0$ hat, und daher ungehindert durch Erhaltungssätze in jedes andere aus einem Teilchen und seinem Antiteilchen bestehende System übergehen kann. Es sind z. B. die folgenden Prozesse möglich: $e^+ + e^- \to 2\gamma$, $e^+ + e^- \to p + \bar{p}$, $e^+ + e^- \to \tau^+ + \tau^-$. Man spricht hier von einer Paarvernichtung im Anfangs- bzw. einer Paarerzeugung im Endzustand. Solche Prozesse werden an den Elektron-Positron-Speicherringen untersucht (vgl. 5.1.4.1.). Bei der Erzeugung schwerer Teilchen wandelt sich die kinetische Energie von Elektron und Positron in Ruhmasse der schweren Teilchen um. Der umgekehrte Effekt liegt bei der Zerstrahlung eines massiven Paares in Photonen vor. Die Mindestenergie, die für die Erzeugung schwerer Teilchen erforderlich ist, bezeichnet man als *Schwellenergie*. Bei dieser Energie werden die neuen Teilchen mit verschwindend kleinem Impuls erzeugt, so daß die Schwellenergie gleich der doppelten Ruhenergie des zu erzeugenden Teilchens ist.

5.1.1.7. Quantenfeldtheorie. Es ist eine der wichtigsten Erkenntnisse unseres Jahrhunderts, daß den Mikroobjekten, d. h. den Elementarteilchen sowohl ein Wellen- als auch ein Teilchenaspekt innewohnt. Dieser zuerst für Photonen konstatierte Sachverhalt fand später vielfältige Bestätigung durch die Beobachtung von Beugungserscheinungen von Elektronen- und Neutronenstrahlen bis hin zu den Interferenzeffekten bei der Ausbreitung neutraler K-Mesonen. Die Quantenfeldtheorie bietet eine konsistente mathematische Formulierung für den Welle-Teilchen-Dualismus, die überdies den Forderungen der relativistischen Invarianz genügt. Jeder Teilchensorte entspricht ein Quantenfeld, das einer relativistischen Feld- oder Wellengleichung genügt. Die Feldvariablen sind jedoch im Gegensatz zur klassischen Physik Operatoren, d. h. nicht miteinander vertauschbare mathematische Größen. Ein Quantenfeld beschreibt immer ein Ensemble von Teilchen der jeweiligen Sorte und von ihren Antiteilchen, die über die Angabe von Impulsen u. a. Quantenzahlen (Spinkomponenten oder inneren Quantenzahlen) hinaus nicht unterscheidbar sind.

Als die grundlegenden theoretischen Resultate der relativistischen Quantenfeldtheorie sind die Vorhersage von Antiteilchen, das Spin-Statistik-Theorem (Teilchen mit halbzahligem Spin sind Fermionen, Teilchen mit ganzzahligem Spin Bosonen) und das TCP-Theorem anzusehen. Letzteres besagt, daß alle Wechselwirkungen unter der Kombination der Symmetrietransformationen Spiegelung, Ladungskonjugation und Zeitumkehr invariant sind. Alle diese Resultate haben eine derart universelle experimentelle Bestätigung gefunden, daß eine theoretische Alternative zur Quantenfeldtheorie ernsthaft nicht zur Diskussion steht.

Um Streuprozesse und Umwandlungen von Elementarteilchen ineinander beschreiben zu können, müssen die Feldgleichungen Wechselwirkungsglieder mit multiplikativ aneinander gekoppelten Feldern enthalten. Solche Theorien können zur Zeit nur mit approximativen Methoden behandelt werden. Eine davon ist die störungstheoretische Berechnung von Wahrscheinlichkeitsamplituden, die auf das anschauliche Bild des Feynman-Graphen führte (vgl. 5.1.1.8.). Die Behandlung der starken Wechselwirkungen macht andere Näherungsmethoden erforderlich. Ein besonderer Aspekt der quantenfeldtheoretischen Methode ist eine gewisse Nichteindeutigkeit, die sich etwa bei der Behandlung der Hadronen äußert. Es ist einerseits völlig legitim, Baryonen oder Mesonen als Quanten entsprechender Felder aufzufassen. Es wäre jedoch unökonomisch, die starken Wechselwirkungen zwischen ihnen mit Hilfe von gekoppelten Feldgleichungen für Baryon- und Mesonenfelder beschreiben zu wollen. Bei einem solchen Vorgehen hätte man es mit einer großen Zahl von Feldern und einer noch

größeren Zahl von Kopplungen zwischen ihnen zu tun. Statt dessen hat sich die Erkenntnis durchgesetzt, daß die starken Wechselwirkungen auf die Kopplungen zwischen fundamentalen Quark- (vgl. 5.1.2.3.) und Gluonfeldern (vgl. 5.1.2.4.) zurückgeführt werden können.

5.1.1.8. Feynman-Graphen haben ihren Ursprung im Formalismus der relativistischen Quantenfeldtheorie und sind eigentlich symbolische Vorschriften zur Berechnung von Wirkungsquerschnitten. Die Feynman-Graphen haben zugleich aber auch einen sehr anschaulichen Gehalt, indem sie unsere Vorstellungen von den Elementarteilchen und ihren Wechselwirkungsprozessen in einer bildhaften Form darstellen. Ein Graph besteht aus Linien und Vertices, d. h. Knotenpunkten. Den Linien sind Teilchen zugeordnet, insbesondere bedeuten Linien mit Pfeil Teilchen mit dem Spin 1/2, eine Wellenlinie das Photon oder die intermediären Bosonen (vgl. 5.1.2.4.), durchbrochene Linien Mesonen. Den Linien ist auch ein Impuls zugeordnet, der in einen Vertex entweder einläuft oder ausläuft. Wenn Pfeilrichtung der Linie und Richtung des zugehörigen Impulses übereinstimmen, beschreibt die Linie ein Teilchen, sonst das Antiteilchen. Die Vertices, in denen sich i. allg. drei oder vier Linien treffen, symbolisieren Wechselwirkungsakte (Abb. 5.1.1–3). An jedem Vertex sind die Erhaltungssätze erfüllt, die für die betreffende Wechselwirkung gelten, d. h., Energie, Impuls und Ladung bleiben in jedem Fall erhalten. Die zwei folgenden Punkte verdienen besondere Beachtung.

Ein Graph beschreibt immer eine Gruppe von Prozessen, die durch die Ersetzung „einlaufendes Teilchen → auslaufendes Antiteilchen" und umgekehrt auseinander hervorgehen. So beschreibt z. B. Abb. 5.1.1–3b den Prozeß $p \to n + \Pi^+$, wenn die Impulse von Neutron und Pion aus dem Vertex laufen, dagegen den Prozeß $p + \bar{n} \to \Pi^+$, wenn die Impulse von Proton und Antineutron in den Vertex einlaufen.

Die inneren Linien eines Graphen, die also zwei Vertices verbinden, stellen virtuelle Teilchen dar. Sie sind im Unterschied zu den durch äußere Linien repräsentierten Teilchen nicht beobachtbar und es gilt auch nicht die übliche Relation zwischen Energie, Impuls und Masse.

Die Wechselwirkungen und Kräfte zwischen reellen Teilchen beruhen auf dem Austausch von virtuellen Teilchen. Die virtuellen Teilchen spielen schon in den einfachsten Prozessen der Quantenelektrodynamik, bei der Elektron-Elektron-Streuung (Abb. 5.1.1–4a), der Compton-Streuung (Abb. 5.1.1–4b) und der Elektron-Positron-Zerstrahlung (Abb. 5.1.1–4c) eine Rolle. Der Austausch eines virtuellen Photons zwischen e^- und e^-, e^- und e^+, e^+ und e^+ erzeugt die Coulomb-Kraft. Die Kernkräfte, d. h. die zwischen den Nukleonen wirkenden Kräfte, haben ihren Ursprung im Austausch von virtuellen Pionen, K-Mesonen usw. (Abb. 5.1.1–5a). Da es sich hier um den Austausch von massiven virtuellen Teilchen handelt, ergeben sich die kurzreichweitigen Kernkräfte.

Jeder Vertex trägt eine Kopplungskonstante, die ein Maß für die Stärke der Wechselwirkung ist. Der Elektron-Photon-Vertex ist proportional zur Elementarladung e. Folglich ist die Kraft zwischen zwei Elektronen bzw. Elektron und Positron proportional zu $(-e)\cdot(-e) = e^2$ bzw. $(-e)(+e) = -e^2$, d. h. einmal abstoßend und das andere Mal anziehend. Als Maß für die Stärke der elektromagnetischen Wechselwirkung benutzt man den dimensionslosen Ausdruck $e^2/4\pi\varepsilon_0\hbar c \approx 1/137$, die sogenannte Sommerfeldsche Feinstrukturkonstante. Der Pion-Nukleon-Vertex und allgemein die Meson-Baryon-Vertices sind nur indirekt und nicht sehr genau bestimmbar. Die zur Feinstrukturkonstante analogen Ausdrücke liegen in der Größenordnung von 10.

Grundsätzlich tragen zu jedem Prozeß Graphen beliebig hoher Ordnung bei (vgl. Abb. 5.1.1–5b). Die sukzessive Berücksichtigung von Graphen immer höherer Ordnung heißt Störungsrechnung. Wegen der Kleinheit der Elektron-Photon-Kopplung stellt der erste Graph den weitaus größten Betrag dar. Das ist der Grund dafür, daß mit großer Genauigkeit die Kraft zwischen Elektronen, und ähnlich die Kraft zwischen Elektron und Proton, durch eine einfache Cou-

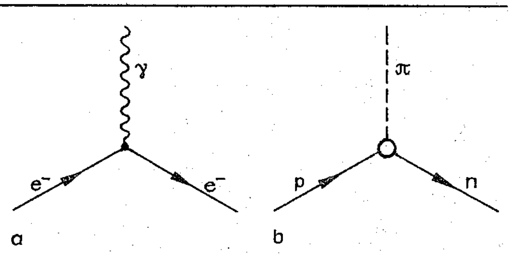


Abb. 5.1.1–3 Fundamentaler Elektron-Photon-Vertex (a), Pion-Nukleon-Vertex (Impulse sind nicht eingezeichnet) (b)

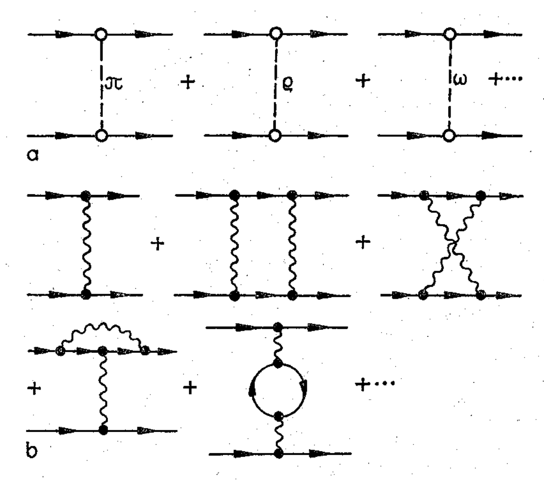
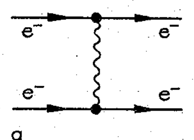


Abb. 5.1.1–5 Beiträge (a) zu den Kräften, die zwischen den Nukleonen wirken, (b) zu den elektromagnetischen Kräften zwischen Elektronen



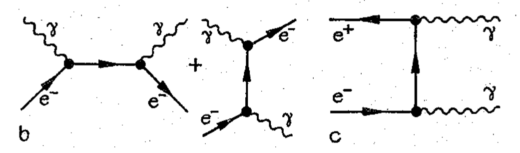


Abb. 5.1.1–4 Feynman-Graphen; (a) zur Streuung von Elektronen aneinander, (b) zur Compton-Streuung an Elektronen, (c) zur Zerstrahlung $e^- + e^+ \to \gamma + \gamma$

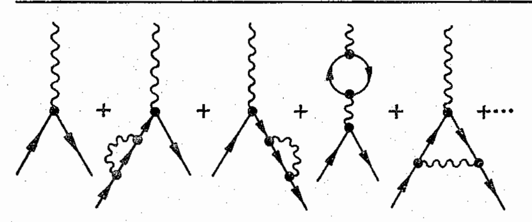


Abb. 5.1.1–6 Beiträge zum Elektron-Photon-Vertex

lomb-Kraft ersetzt werden kann. Diesem glücklichen Umstand ist es zu verdanken, daß die Quantentheorie weitgehende Einsichten in die Gesetzmäßigkeiten der Atomhülle gewinnen konnte. Die höheren Graphen sind in der Quantenelektrodynamik berechnet worden. Sie führen auf die Strahlungskorrekturen der Kräfte, die sich in kleinen Verschiebungen der Energieniveaus der Atomhülle auswirken. Auch zum magnetischen Moment μ des Elektrons tragen Graphen beliebig hoher Ordnung (Abb. 5.1.1–6) bei. Der erste Graph, also der fundamentale Vertex, liefert das normale magnetische Moment, die übrigen Graphen den anormalen Beitrag. Die berechneten Strahlungskorrekturen stimmen ausgezeichnet mit den experimentellen Fakten überein.

Im Unterschied dazu ist die starke Wechselwirkung zwischen Nukleonen wegen der Größe der Kopplungskonstante nicht auf dem Wege der Störungsrechnung zu behandeln. Es trägt in jedem Fall eine quantitativ nicht überschaubare Reihe von Graphen bei (vgl. Abb. 5.1.1–5a). Das gilt gleichermaßen, ob nun die Kernkräfte auf den Austausch von Mesonen zurückgeführt werden (vgl. Abb. 5.1.1–5a) oder, was vollkommen äquivalent ist, auf den Austausch von Gluonen und Quark-Antiquark-Paaren im Rahmen der Quantenchromodynamik. Daher konnten die Kernkräfte bisher nicht in ähnlicher Weise berechnet werden wie die

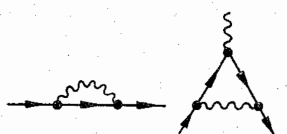


Abb. 5.1.1–7 Beispiele für divergierende Graphen

elektromagnetischen Kräfte. Der Methode der Feynman-Graphen haftet eine prinzipielle Schwierigkeit an. Gewisse Graphen mit geschlossenen Linienzügen (Abb. 5.1.1–7) ergeben bei der Berechnung ein unendliches Resultat. Zur Lösung dieses Problems ist die mathematisch sehr anspruchsvolle Methode der Renormierung entwickelt worden. Bei diesem Verfahren werden die experimentell gemessenen Werte von Kopplungskonstante und Elektronenmasse in der Theorie als Differenzen divergierender Größen dargestellt. Berechenbar sind nur Strahlungskorrekturen und Wirkungsquerschnitte.

5.1.2. Leptonen, Hadronen, Quarks und intermediäre Bosonen

Eine natürliche Einteilung der Teilchen beruht darauf, daß sie in unterschiedlichem Maße den verschiedenen Typen von Wechselwirkungen unterliegen (vgl. 5.1.3.). So lassen sich alle bisher beobachteten Teilchen in die großen Gruppen der Leptonen, der Hadronen und der intermediären Bosonen einordnen. In den Tab. 5.1.2–1 und 5.1.2–2 sind die relativ langlebigen Leptonen und Hadronen mit ihren Quantenzahlen zusammengestellt.

5.1.2.1. Leptonen. In diese Gruppe gehören das Elektron e^- und das Elektron-Neutrino ν_e, das Myon μ^- und das Myon-Neutrino ν_μ sowie das schwere Lepton τ. Der Umstand, daß Elektron und Myon ($M_e = 0{,}511\,\mathrm{MeV}/c^2$, $M_\mu = 106\,\mathrm{MeV}/c^2$) und auch die Neutrinos sehr leichte Teilchen sind, hat zu der Bezeichnung Leptonen, d. h. leichte Teilchen, Anlaß gegeben. Die Eigenart der Leptonen besteht darin, nicht der starken Wechselwirkung zu unterliegen (vgl. 5.1.3.2.). Das Myon zerfällt im Prozeß $\mu^- \to e^- + \bar{\nu}_e + \nu_\mu$ mit einer Lebensdauer von $2{,}2 \cdot 10^{-6}$ s. Das Neutrino ist ursprünglich nur hypothetisch eingeführt worden, um die scheinbare Verletzung der Energieerhaltung beim β-Zerfall $n \to p + e^- + \bar{\nu}_e$ zu erklären. Da das Neutrino ungeladen ist, unterliegt es ausschließlich der schwachen Wechselwirkung mit dementsprechend kleinen Wirkungsquerschnitten, und kann daher nur äußerst schwer nachgewiesen werden. Bezüglich der Masse des Elektron-Neutrinos liegen bisher nur die Ergebnisse eines einzigen Präzisionsexperiments vor, das die Neutrinomasse aus dem Energiespektrum der beim β-Zerfall des Tritiums ausgesandten Elektronen bestimmt: $20\,\mathrm{eV}/c^2 < M_{\nu_e} < 40\,\mathrm{eV}/c^2$. Selbstverständlich bedarf diese Aussage einer Bestätigung durch andere unabhängige Experimente.

Der Nachweis der Neutrinos erfolgt in der Regel aufgrund der durch sie initiierten Reaktionen. Das sind in der Nachweisreaktion der Elektron-Neutrinos Kernumwandlungen, z. B. $\nu_e + {}^{37}\mathrm{Cl} \to {}^{37}\mathrm{Ar} + e^-$, deren Endprodukt quantitativ erfaßt wird. Antineutrinos lassen sich mit Hilfe der Reaktion $\bar{\nu}_e + p \to n + e^+$ nachweisen. Beobachtet werden in diesem Falle die Photonen, die bei den nachfolgenden Einfangprozessen von n und e^+ entstehen. Myon-Neutrinos werden durch die in den Reaktionen $\nu_\mu + \text{Nukleon} \to \mu^- + \ldots$ und $\bar{\nu}_\mu + \text{Nukleon} \to \mu^+ + \ldots$ gebildeten μ^- und μ^+ angezeigt. Der experimentelle Beleg für die Verschiedenheit von ν_e und ν_μ besteht darin, daß die beim Pionzerfall entstandenen Neu-

Tab. 5.1.2–1 Zusammenstellung der Leptonen

	M in MeV/c^2	τ in s	Q in $\|e\|$	I in $\hbar$	L_e	L_μ	L_τ	Zerfälle (u. a.)
Elektron e^-	0,511	∞	−1	1/2	+1	0	0	
Myon μ^-	105,7	$2{,}20 \cdot 10^{-6}$	−1	1/2	0	+1	0	$e^-\ \bar{\nu}_e\ \nu_\mu$
τ-Lepton τ^-	1782	$3 \cdot 10^{-13}$	−1	1/2	0	0	+1	$\mu^-\ \bar{\nu}_\mu\ \nu_\tau$
Elektronneutrino ν_e	$(20\ldots40) \cdot 10^{-6}$	∞	0	1/2	+1	0	0	
Myonneutrino ν_μ	< 0,5	?	0	1/2	0	+1	0	
τ-Neutrino	< 143	?	0	1/2	0	0	+1	

Tab. 5.1.2–2 Zusammenstellung der relativ langlebigen Hadronen

	M in GeV/c^2	τ in s	Q in |e|	I in $\hbar$	P	I	I_3	Y	S	C	Zerfälle (u. a.)
Mesonen											
Pion Π°	0,135	$8{,}3 \cdot 10^{-17}$	0	0	−1	1	0	0	0	0	2γ
Pion Π^+	0,139	$2{,}6 \cdot 10^{-8}$	+1	0	−1	1	+1	0	0	0	$\mu^+ \nu_\mu$
η-Meson η	0,549	$\Gamma = 0{,}8$ KeV	0	0	−1	0	0	0	0	0	2γ, $3\pi^\circ$
Kaon K^+	0,494	$1{,}24 \cdot 10^{-8}$	+1	0	−1	1/2	+1/2	−1	+1	0	$\mu^+ \nu_\mu$, $\pi^+ \pi^\circ$
Kaon K°	0,498	50% $8{,}9 \cdot 10^{-11}$	0	0	−1	1/2	−1/2	+1	+1	0	
Kaon $\bar{K}^\circ$	0,498	50% $5{,}2 \cdot 10^{-8}$	0	0	−1	1/2	+1/2	−1	+1	0	
D-Meson D^+	1,869	$9 \cdot 10^{-13}$	+1	0	−1	1/2	1/2		0	+1	$\bar{K}^\circ \pi^+ \pi^\circ$
D-Meson D°	1,863	$4 \cdot 10^{-13}$	0	0	−1	1/2	−1/2		0	+1	$K^- \pi^+ \pi^\circ$
F-Meson F^+	1,975	$3 \cdot 10^{-13}$	+1	0	−1	0	0		+1	+1	
B-Meson B^+	5 271	$\sim 10^{-12}$	+1	0	−1	1/2	+1/2		0	0	
B-Meson B°	5 274	$\sim 10^{-12}$	0	0	−1	1/2	−1/2		0	0	
Baryonen											
Proton p	0,9383	$> 10^{31}$ Jahre	+1	1/2	+1	1/2	+1/2	+1	0	0	
Neutron n	0,9396	900	0	1/2	+1	1/2	−1/2	+1	0	0	$p e^- \bar{\nu}_e$
Lambda Λ	1,116	$2{,}6 \cdot 10^{-10}$	0	1/2	+1	0	0	0	−1	0	$p\pi^-$, $n\pi^\circ$
Sigma Σ^+	1,189	$0{,}8 \cdot 10^{-10}$	+1	1/2	+1	1	+1	0	−1	0	$p\pi^\circ$, $n\pi^+$
Sigma Σ°	1,192	$5{,}8 \cdot 10^{-20}$	0	1/2	+1	1	0	0	−1	0	$\Lambda \gamma$
Sigma Σ^-	1,197	$1{,}5 \cdot 10^{-10}$	−1	1/2	+1	1	−1	0	−1	0	$n \pi^-$
Xi Ξ°	1,315	$2{,}9 \cdot 10^{-10}$	0	1/2	+1	1/2	+1/2	−1	−2	0	$\Lambda \pi^\circ$
Xi Ξ^{-1}	1,321	$1{,}6 \cdot 10^{-10}$	−1	1/2	+1	1/2	−1/2	−1	−2	0	$\Lambda \pi^-$
Omega Ω^-	1,671	$0{,}9 \cdot 10^{-10}$	−1	3/2	+1	0	0	−2	−3	0	ΛK^-, $\Xi^\circ \pi^-$
$\Lambda_c{}^+$	2,282	$\sim 2 \cdot 10^{-13}$	+1	1/2	+1	0	0		0	+1	
A^+	2,460	$\sim 5 \cdot 10^{-13}$	+1	1/2	+1	1/2	+1/2		−1	+1	

trinos bei ihrer Reaktion mit Nukleonen immer nur Myonen, jedoch niemals Elektronen erzeugen; es handelt sich also um Myon-Neutrinos. Wenn sich die Neutrinomassen als von Null verschieden erweisen sollten, könnten sich auch die Eigenschaften der Neutrinos als verwickelter herausstellen, als bisher angenommen worden ist. Es bestände dann die Möglichkeit, daß man unterscheiden müßte zwischen den physikalischen, d.h. als freie Teilchen existierenden Neutrinos, und den in den Wechselwirkungsprozessen auftretenden ν_e, $\bar{\nu}_e$, ν_μ und $\bar{\nu}_\mu$. Die Erhaltungssätze für L_e und L_μ hätten nur für die Wechselwirkungsakte Geltung, wogegen die freien Neutrinos als quantentheoretische Überlagerung aus ν_e, $\bar{\nu}_e$, ν_μ usw. aufzufassen wären. Nach diesen sogenannten Neutrinooszillationen wird gegenwärtig intensiv gesucht. Heute ist es möglich, mit allen Arten von Neutrinos Streuexperimente durchzuführen. Diese Neutrinos stammen entweder aus dem β-Zerfall der in Kernreaktoren erzeugten Neutronen oder aus dem Zerfall von Kaonen oder Pionen gemäß $K^- \to \mu^- + \bar{\nu}_\mu$ oder $K^+ \to \mu^+ + \nu_\mu$.

Im Unterschied zu den meisten Elementarteilchen ist das Elektron-Neutrino eine Komponente der natürlichen irdischen Umwelt. Die im Innern der Sonne u.a. Sterne ablaufenden Kernreaktionen sind Quelle eines Stromes von Neutrinos ν_e, die praktisch ungeschwächt sowohl die Atmosphäre als auch den Erdkörper durchdringen. Gegenwärtig werden große Anstrengungen unternommen, um diesen Teilchenstrom quantitativ zu erfassen. Ein solches Großexperiment wird im Kaukasus (UdSSR) vorbereitet, mehr als 1 500 m unter Tage, um jegliche Störung durch Teilchen der Höhenstrahlung auszuschließen. Die Menge des in riesigen Kavernen exponierten Detektormateriales beträgt einige 1 000 t. Die Stabilität und die kleinen Wirkungsquerschnitte führen zu dem Schluß, das selbst in kosmischen Dimensionen Neutrinos kaum absorbiert werden. Man hat überdies mit einer Neutrinountergrundstrahlung zu rechnen, die ebenso wie die elektromagnetische Untergrundstrahlung als ein Relikt aus einer Frühphase der kosmischen Entwicklung übriggeblieben ist.

Eine der bedeutungsvollsten Entdeckungen der letzten Jahre ist die des schweren Leptons τ^- und seines Antiteilchens τ^+. Die Familie der Leptonen hatte seit dreißig Jahren für abgeschlossen gegolten. Gesichert ist, daß τ ein einfach geladenes Teilchen mit dem Spin $J = 1/2$ und einer Masse von $\approx 1{,}8$ GeV/c^2 ist. Das neue Lepton wurde an den Elektron-Positron-Speicherringen dadurch entdeckt, daß Endzustände beobachtet wurden, die nur aus μ^-, e^+ oder μ^+, e^- und zusätzlich aus Neutrinos bestehen. Man erklärt sich dies so, daß zuerst ein τ-Paar in der Reaktion $e^+ + e^- \to \tau^+ + \tau^-$ erzeugt worden ist, dessen Partner anschließend in den Reaktionen $\tau^+ \to e^+ + \nu_e + \bar{\nu}_\tau$ und $\tau^- \to e^- + \bar{\nu}_e + \nu_\tau$ zerfallen. Es kann als gesichert gelten, daß zum τ-Lepton ein eigenes τ-Neutrino gehört. Aus den beobachteten τ-Zerfällen ergibt sich die Abschätzung $M\nu_\tau < 143$ MeV/c^2. Der Umstand, daß τ offenbar nicht der starken Wechselwirkung unterliegt, rechtfertigt die Bezeichnung Lepton, obwohl seine Masse groß ist. In einigen der Zerfallskanäle von τ konnte man erstmalig Hadronen als Zerfallsprodukte eines Leptons beobachten, z. B. das Pion in dem Zerfall $\tau^- \to \Pi^- + \nu_\tau$.

5.1.2.2. Hadronen. Die charakteristische Eigenschaft von Hadronen ist es, allen Typen von Wechselwirkungen zu unterliegen. Die Hadronen mit halbzahligem Spin heißen *Baryonen*, diejenigen mit ganzzahligem Spin *Mesonen*. Sehr kurzlebige Hadronen, wie etwa die zur Zeit bekannten Mesonen mit dem Spin 1 und höher werden *Resonanzen* genannt. Sie zeigen sich nur als Maxima in der Energieab-

hängigkeit der Wirkungsquerschnitte. Die bekanntesten Baryonen sind Proton und Neutron, die auch als *Nukleonen* bezeichnet werden, da sie die Bausteine der Atomkerne sind. Das Proton ist stabil, das Neutron mit einer mittleren Lebensdauer von 900 s quasistabil. Wenn das Neutron sich jedoch in stabilen Atomkernen oder in Neutronensternen befindet, kann der Zerfall energetisch unmöglich sein. Das einfachste Beispiel bietet der Deuteriumkern, der ein gebundenes System aus Proton und Neutron ist. Wenn das Neutron durch β-Zerfall in ein Proton übergehen wollte, müßte es zugleich den Kern verlassen, da es keinen aus zwei Protonen bestehenden Kern gibt. Dafür aber reicht die ursprüngliche Energie nicht aus.

Alle Hadronen, außer Proton und Neutron, müssen durch Stoßprozesse erzeugt werden. Es dominiert hierbei selbstverständlich die starke Wechselwirkung, die die größten Wirkungsquerschnitte ergibt. Die in den Stößen erzeugten Teilchen darf man sich nicht als Bestandteile der aufeinandertreffenden Stoßpartner vorstellen. Bei den Erzeugungsprozessen gelten Gesetzmäßigkeiten, die man aufgrund des Erhaltungssatzes für die Hyperladung verstehen kann. Danach ist zum Beispiel der Prozeß $p + p \rightarrow p + \Sigma^+ + K^0$ möglich, nicht aber $p + p \rightarrow \Sigma^+ + \Sigma^0 + \Pi^0$. Die historische Entwicklung verlief so, daß die Quantenzahl Y eingeführt worden ist, um die experimentell beobachteten Regeln zu begründen.

Die Zerfallsprozesse der Hadronen sind i. allg. komplizierter als die der Leptonen. Von entscheidender Bedeutung ist, nach welchem Typ von Wechselwirkung der Zerfall abläuft. Die Benutzung des Begriffs Wechselwirkung in diesem Zusammenhang läßt sich dadurch veranschaulichen, daß der umgekehrte Prozeß ein Wechselwirkungsakt im eigentlichen Sinne ist. In der Regel ist die Lebensdauer um so kürzer, je stärker die Zerfallswechselwirkung ist. Zerfallsprozesse, die unter Erhaltung von Isospin und Hyperladung ablaufen, wie etwa der Zerfall der Baryonresonanz Σ^{*+} ($\Sigma^{*+} \rightarrow \Sigma^+ + \Pi^0$), gehorchen der starken Wechselwirkung und beschränken die Lebensdauer auf die Größenordnung 10^{-23} s. Wenn Isospin und Hyperladung nicht erhalten bleiben, z. B. in $\Sigma^+ \rightarrow n + \Pi^+$, so erfolgt der Zerfall nach der schwachen Wechselwirkung, und die Lebensdauer ist von der Größenordnung 10^{-10} s. Ähnliches gilt für die Baryonen Σ^-, Λ, Ξ^0, Ξ^- und die Mesonen K^+ und K^-.

Eine besonders interessante Situation liegt bei den neutralen Kaonen vor. Sie werden in einem Prozeß der starken Wechselwirkung als Teilchen K^0 oder $\bar{K}^0$ erzeugt, denen die Hyperladung $Y = +1$ bzw. $Y = -1$ zugeordnet ist. Nachdem die Kaonen das Gebiet verlassen haben, wo sie erzeugt worden sind, kommt es i. allg. zu keinem weiteren Akt der starken Wechselwirkung; allein die schwache Wechselwirkung ist maßgeblich für die weitere zeitliche Entwicklung. Als beobachtbare Teilchen existieren sie daher nur in den Zuständen K^0_S und K^0_L, die als quantentheoretische Superpositionen der beiden ursprünglichen Zustände K^0 und $\bar{K}^0$ aufzufassen sind, und denen kein bestimmter Wert von Y zugeordnet werden kann. Dieser Sachverhalt läßt sich auch so ausdrücken, daß 50 % aller K^0 (oder $\bar{K}^0$) als K^0_S zerfallen, und die restlichen 50 % als K^0_L. Beim Zerfall des längerlebigen Zustands K^0_L ist der bislang einzige Prozeß entdeckt worden, der unter CP-Verletzung abläuft. Hier bedeutet CP die Kombination aus Raumspiegelung und Ladungskonjugation, die Teilchen und Antiteilchen ineinander überführt. Für den Zerfall der neutralen K-Mesonen ist CP keine Symmetrieoperation.

Einen wesentlichen Zuwachs hat die Familie der Hadronen durch die sogenannten neuen Teilchen erfahren. Sie zeigten sich erstmalig als Resonanzen im Wirkungsquerschnitt der Elektron-Positron-Vernichtung (vgl. 5.1.4.1). Diese Entdeckung hat darum so große Überraschung ausgelöst, weil überhaupt nicht zu erwarten war, daß im Massenbereich oberhalb von 2 GeV/c^2 noch derartig scharfe Resonanzen existieren. Die Resonanzbreite von ψ (3,097 GeV/c^2) ist 0,06 MeV/c^2, und auch die Breite von ψ' (3,685 GeV/c^2) ist mit 0,22 MeV/c^2 noch sehr klein. Die neuen Teilchen ψ, ψ' und neuerdings Y und Y' sind in Vielem mit den schon länger bekannten Resonanzen ϱ^0, ω und Φ zu vergleichen (Tab. 5.1.2–3). Alle treten sie im e^+–e^--Kanal auf und haben die Quantenzahlen $J = 1$ und $P = -1$. Ihre ungewöhnliche Schärfe findet im Quarkmodell eine hinreichende Erklärung. Wie es sich seither gezeigt hat, waren diese Teilchen im e^+–e^--Kanal Vorboten von Gruppen neuer Teilchen (den Mesonen D^0, D^+, F^+, B^0 und den Baryonen Λ_c^+, A^+), in denen sich die Quantenzahlen „Charm" und „Beauty" als neue Freiheitsgrade der Hadronen manifestieren. Jedoch bietet das Quarkmodell eine weitaus anschaulichere Deutung hierfür.

Tab. 5.1.2–3 Resonanzen im e^+e^--Kanal

Symbol	M in GeV/c^2	Breite in MeV/c^2
ϱ^0	0,770	150
ω	0,783	10
Φ	1,020	4
Ψ	3,097	0,06
Ψ'	3,685	0,22
Y	9,46	0,05
Y'	10,02	0,03

5.1.2.3. Quarks. Die Ergebnisse langjähriger experimenteller und theoretischer Untersuchungen der Hadronen und ihrer Wechselwirkungen haben im Quarkmodell ihre zusammenfassende Deutung gefunden: Baryonen bestehen im Sinne quantentheoretischer gebundener Zustände aus drei Quarks, Mesonen dagegen aus Quark und Antiquark. Deshalb müssen die Quarks halbzahligen Spin und gebrochene Werte von elektrischer Ladung und Baryonzahl tragen (Tab. 5.1.2–4).

Tab. 5.1.2–4 Quantenzahlen der Quarks

Quark	B	Q	I	I_3	Y	S	C	Beauty
u	1/3	2/3	1/2	+1/2	1/3	0	0	0
d	1/3	−1/3	1/2	−1/2	1/3	0	0	0
s	1/3	−1/3	0	0	−2/3	−1	0	0
c	1/3	2/3	0	0		0	+1	0
b	1/3	−1/3	0	0		0	0	+1
t	1/3	2/3	0	0		0	0	0

Der Ursprung des Quarkmodells liegt in den Vorstellungen der unitären Symmetrie, die als ein Ordnungsprinzip zum Verständnis der leichteren Hadronen diente. Es zeigte sich, daß die vor der Entdeckung der „neuen Teilchen" bekannten Baryonen und Mesonen sich in Achter- und Zehnermultipletts einordnen ließen, wobei auch Isospin und Hyperladung der Teilchen in allen Fällen den von der unitären Symmetrie geforderten Werten entsprachen (Abb. 5.1.2–5).

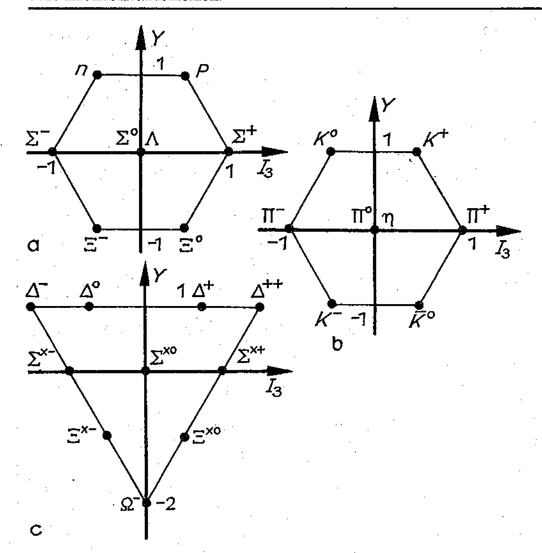


Abb. 5.1.2–5 Multipletts der Hadronen; (a) Oktett der Baryonen mit Spin 1/2, (b) Oktett der Mesonen mit Spin 0, (c) Dekuplett der Baryonen mit Spin 3/2

Die unitäre Symmetrie ist allerdings nur in gebrochener Form realisiert, da sich die Mitglieder eines Multipletts hinsichtlich der Masse unterscheiden. Es gelang jedoch, die Struktur der Symmetriebrechung aufzuklären, Massenformeln herzuleiten und auf diesem Wege das zu jener Zeit noch nicht bekannte Baryon Ω^- vorherzusagen. Die unitäre Symmetrie läßt als kleinste Multipletts zwei Tripletts zu, aus denen sich in mathematisch eindeutigem Sinne alle übrigen Multipletts erzeugen lassen. Damit war die Existenzmöglichkeit für die Quarks *u*, *d*, *s* und ihre Antiteilchen theoretisch eröffnet (Abb. 5.1.2–6). Eine unmittelbare Begründung hat das Quarkmodell aus den Experimenten zur tief-inelastischen Lepton-Hadron-Streuung (vgl. 5.1.4.2.) erhalten.

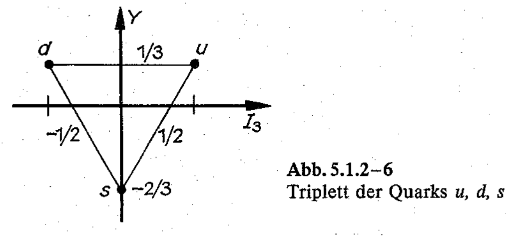


Abb. 5.1.2–6 Triplett der Quarks *u*, *d*, *s*

Die Entdeckung der neuen Teilchen hat die Idee der unitären Symmetrie in den Hintergrund gedrängt; die Unterschiede in den Massen sind derartig groß, daß z.B. die Einordnung der Mesonen Π, K, D und B in ein einziges großes Multiplett ihren Sinn verlieren würde. Geblieben jedoch ist die Vorstellung von Quarks. Jede Entdeckung neuer Freiheitsgrade der Hadronen (Charm, Beauty, ...) sieht man als Manifestation neuer Quarks an. In diesem Sinne ist auch die 1984 erfolgte Entdeckung des *t*-Quarks zu verstehen. Freie Quarks wurden bisher nicht gefunden. Die Quarkstruktur einiger Hadronen ist in Tab. 5.1.2–7 aufgeführt. Allerdings ist das Quarkmodell viel komplizierter, als es dieser grobe Abriß zeigen könnte. Gewichtige theoretische

Tab. 5.1.2–7 Quarkstruktur einiger Hadronen

p	n	Σ^+	Ω^-	Λ_c^+	Π^+	K^+	Ψ	D^+	Y	B^+
uud	*udd*	*uus*	*sss*	*udc*	$u\bar{d}$	$u\bar{s}$	$c\bar{c}$	$c\bar{d}$	$b\bar{b}$	$u\bar{b}$

Argumente, u. a. die Antisymmetrie der quantenmechanischen Wellenfunktion der Baryonen, fordern, daß die Quarks durch eine zusätzliche Quantenzahl gekennzeichnet werden, die den Namen „Color" erhalten hat. Jeder der Quarks, also etwa *u*, kann danach in drei verschiedenen Colorzuständen existieren, die ein Colortriplett bilden. Im Unterschied dazu sollten die Hadronen Colorsingletts sein, also nicht in verschiedenen Colorzuständen auftreten. Die starken Wechselwirkungen zwischen den Quarks haben ihren Ursprung im Austausch von Gluonen (vgl. 5.1.2.4.), die ebenfalls Colorladungen tragen. Die Colorsymmetrie ist im Gegensatz zur ursprünglichen unitären Symmetrie, die die *u*-, *d*- und *s*-Quarks vereinte, als ungebrochen aufzufassen; insbesondere sind die Massenwerte innerhalb von Colormultipletts gleich.

Detaillierte Untersuchungen der Quantenchromodynamik, der Feldtheorie der Quarks und Gluonen, haben Argumente für die Nichtexistenz freier Quarkteilchen erbracht, in der Fachsprache „Quark Confinement" genannt.

So können auch die Massen, die den Quarks zugeordnet werden, keine unmittelbare physikalische Bedeutung haben. Wenn man die Quarkmassen aus den Massen der Teilchen im e^+–e^--Kanal (vgl. Tab. 5.1.2–3) abstrahiert, was im Falle der schweren Teilchen gerechtfertigt ist, ergeben sich die folgenden Werte: $M_s \approx 150\,\mathrm{MeV}/c^2$, $M_c \approx 1{,}6\,\mathrm{GeV}/c^2$, $M_b \approx 5\,\mathrm{GeV}/c^2$ und $M_t \approx 30$–$50\,\mathrm{GeV}/c^2$.

5.1.2.4 Intermediäre Bosonen. Nach der heutigen Einsicht werden alle Wechselwirkungen zwischen Elementarteilchen durch den Austausch einer besonderen Gruppe von Teilchen vermittelt, den intermediären Bosonen. Ihnen allen ist der Spin 1 gemeinsam, sie unterscheiden sich jedoch hinsichtlich der Masse und bestimmter Ladungsquantenzahlen (Tab. 5.1.2–8). Die intermediären Bosonen $W^\pm$ und Z^0 sind 1983 an den $p\bar{p}$-Speicherringen des Europäischen Kernforschungszentrums in Genf nachgewiesen worden. Die Massen von W^+, W^- (als Teilchen-Antiteilchen-Paar haben

Tab. 5.1.2–8 Intermediäre Bosonen

	J	*M* in GeV/c^2	*Q* in $\lvert e \rvert$	Colour-Ladung
Photon γ	1	0	0	0
W^+	1	81	+1	0
Z^0	1	94	0	0
Gluon	1	0	0	vorhanden

beide dieselbe Masse) und von Z^0 stimmen innerhalb der Meßfehlergrenzen mit den durch die einheitliche Theorie (vgl. 5.1.3.6.) vorhergesagten Werten überein. Was die Gluonen anbelangt, so gibt es zahlreiche experimentelle Hinweise auf die Existenz eines Oktetts von Gluonenfeldern; wie im Falle der Quarks erwartet man jedoch nicht, daß die Gluonen als physikalisch nachweisbare Teilchen existieren. Der in der Tab. angegebene Massenwert Null ist in diesem Falle als der Wert des in den Feldgleichungen auftretenden Massenparameters zu verstehen. Übrigens be-

stehen begründete Erwartungen, daß sich einige der Resonanzen nur als gebundene Systeme aus Gluonen verstehen lassen.
Formal sind die intermediären Bosonen durch die Eigenschaft charakterisiert, Feldquanten sogenannter Eichfelder zu sein. Das elektromagnetische Feld ist das seit langem bekannte und auch das einfachste Beispiel eines Eichfeldes. Das strukturbestimmende Prinzip der Eichfelder ist die Invarianz der Feldgleichungen gegenüber lokalen, d.h. ortsabhängigen Symmetrietransformationen; dadurch werden die Kopplungen der Eichfelder und damit auch ihre Wechselwirkungen eindeutig festgelegt. Es ergibt sich insbesondere, daß die Eichfelder direkte Kopplungen miteinander haben (vgl. Abb. 5.1.3–3), falls sie ladungsartige Quantenzahlen tragen. Solche Kopplungen fehlen beim Photon.

5.1.2.5. Die Familien der fundamentalen Fermionen. Die als fundamental anzusehenden Felder der Leptonen und Quarks weisen in ihren Eigenschaften eine Periodizität auf, die es nahelegt, sie in drei Gruppen, die sogenannten Familien, zu ordnen (vgl. Tab. 5.1.2–9, aber auch Tab. 5.1.3–9). Die Kopplungen der Leptonen an die intermediären Bosonen γ, W^+ und Z^0 sind von Familie zu Familie dieselben, dasselbe gilt für die Kopplungen der sechs Quarktripletts an die Gluonen.

Tab. 5.1.2–9 Familien der fundamentalen Fermionen

Q	0	−1	2/3	−1/3
1. Familie	ν_e	e^-	u	d
2. Familie	ν_μ	μ^-	c	s
3. Familie	ν_τ	τ^-	t	b

Abgesehen von den Neutrinos, über deren Massen noch nichts Endgültiges ausgesagt werden kann, wachsen die Massen einander entsprechender Familienmitglieder mit steigender Familiennummer an. Zur Zeit sind drei vollständige Familien bekannt, und es gibt keine überzeugenden Gründe, die weitere Familien verbieten würden. Von der Tatsache der Familienstruktur nehmen verschiedene theoretische Spekulationen ihren Ausgang (vgl. 5.1.3.7. und 5.1.5.3.).

5.1.3. Hierarchie der Wechselwirkungen

Die Wechselwirkungen zwischen den Elementarteilchen lassen sich zuerst einmal nach ihrer Stärke ordnen, und zum anderen nach dem Ausmaß, in dem die Erhaltungssätze respektiert werden. Beide Gesichtspunkte führen auf ein- und dasselbe Schema (Tab. 5.1.3–1).

Tab. 5.1.3–1 Überblick über die Wechselwirkungen zwischen Elementarteilchen

	Stärke	Reichweite in cm	gute Quantenzahlen
stark	10	10^{-13}	$B, L_e \dots L_\tau, Q, I, I_3, S, C, P$
elektromagnetisch	10^{-2}	∞	$B, L_e \dots L_\tau, Q, I_3, S, C, P$
schwach	10^{-14}	10^{-16}	$B, L_e \dots L_\tau, Q$

5.1.3.1. Erhaltungssätze. Eine physikalische Größe genügt einem Erhaltungssatz, wenn bei jeder Reaktion ihre Werte in Anfangs- und Endzustand übereinstimmen.
Der Erhaltungssatz für Energie und Impuls wird als absolut gültig angesehen, da noch in keinem Falle eine Verletzung beobachtet worden ist. Energie und Impuls eines Systems aus mehreren Teilchen setzen sich additiv aus den Größen der einzelnen Teilchen zusammen. Nach dem relativistischen Ausdruck für die Energie eines Teilchens mit Masse M und Impuls p besteht zwischen Energie und Impuls die Beziehung $E^2 - p^2c^2 = M^2c^4$. Energie- und Impulssatz werden benutzt, um Resonanzen anhand ihrer Zerfallsprodukte zu identifizieren. Als Beispiel soll der Zerfall $\varrho^0 \rightarrow \Pi^+ + \Pi^-$ dienen. Die Impulse p_1 und p_2 der Pionen werden gemessen; daraus können dann die Energien E_1 und E_2 berechnet werden, sobald die Teilchen als Pionen identifiziert worden sind. Der Ausgangszustand hat die Energie $E = E_1 + E_2$ und den Impuls $p = p_1 + p_2$, woraus sich die Masse zu $M^2 = (E^2 - p^2c^2)/c^4$ ergibt. Das Kennzeichen einer Resonanz ist, daß sich die so berechneten Massen mit nicht allzu großer Streuung um einen mittleren Wert, in unserem Beispiel um die Masse M_ϱ, gruppieren.
Der Erhaltungssatz für den Drehimpuls sagt aus, daß im Ruhsystem der Resonanz der Drehimpuls der Zerfallsprodukte (im Falle der spinlosen Pionen ihr resultierender Bahndrehimpuls) gleich dem Spin der Resonanz ist. Erschwert wird die Anwendung dadurch, daß der Bahndrehimpuls nicht unmittelbar in jedem einzelnen Prozeß meßbar ist. Nach der Quantentheorie läßt er sich aber aus der Winkelverteilung der Reaktionsprodukte bestimmen. Auf diese Weise wird der Spin von Resonanzen und solchen Teilchen ermittelt, die nicht im freien Flug beobachtet werden können.
Die Erhaltung der Parität bedeutet in anschaulicher und experimentell direkt prüfbarer Fassung: Jeder Prozeß, der einer Spiegelungstransformation unterworfen wird, d. h. in einem Spiegel betrachtet wird, läuft mit denselben Wirkungsquerschnitten ab, wie der ursprüngliche Prozeß. Bei schwachen Wechselwirkungen gilt die Paritätserhaltung nicht, d. h., diese Prozesse sind nicht spiegelungsinvariant.
Für die elektrische Ladung Q gilt ein uneingeschränkt gültiger Erhaltungssatz. Auch die Leptonenzahlen L_e, L_μ und L_τ bleiben mit großer Wahrscheinlichkeit erhalten. Die Erhaltung der Baryonenzahl B kann für alle Streuprozesse als praktisch gültig angesehen werden. Nach einer Verletzung dieses Erhaltungssatzes wird in den Experimenten zum Protonzerfall (vgl. 5.1.4.3.) gesucht; es handelt sich hier jedoch um sehr unwahrscheinliche, d. h. sehr seltene Prozesse. Hyperladung Y und Charm C bleiben bei starken und elektromagnetischen Wechselwirkungen erhalten, aber nicht bei schwachen. Die Werte B, L_e, L_μ, L_τ, Q, Y und C, die einem Mehrteilchensystem zugeordnet sind, setzen sich einfach additiv aus den Werten der einzelnen Teilchen zusammen. Als Beispiel soll das Baryon Ω^- mit $Y = -2$ dienen. Ein Zerfall unter Einfluß der starken Wechselwirkungen könnte wegen des Erhaltungssatzes für Y nur in $\Xi^0 + K^-$ oder $\Sigma^+ + K^- + K^-$ erfolgen. Beides ist wegen des Erhaltungssatzes für die Energie nicht möglich. Folglich zerfällt Ω^- unter Einfluß der schwachen Wechselwirkung und ist daher langlebig.
Der Isospin bleibt bei starken Wechselwirkungen erhalten. Die Addition der Isospins der einzelnen Teilchen hat allerdings analog zu den quantentheoretischen Regeln für die Addition von Drehimpulsen zu erfolgen. So ist z. B. der Isospin des aus Pion ($I = 1$) und Nukleon ($I = 1/2$) beste-

henden Systems entweder 1 + 1/2 = 3/2 oder 1 − 1/2 = 1/2, d.h., man findet Resonanzen mit $I = 3/2$ oder $I = 1/2$. Aus der Erhaltung des Isospins lassen sich auch Aussagen über den Zerfall von Resonanzen machen, z.B., daß die Zerfälle $\Delta^0 \rightarrow \Pi^- + p$ und $\Delta^0 \rightarrow \Pi^0 + n$ mit der relativen Häufigkeit 1:2 ablaufen.

Quantenzahlen, für die ein Erhaltungssatz gilt, heißen gute Quantenzahlen. Ganz allgemein entspricht jedem Erhaltungssatz eine Symmetrie oder Transformationsinvarianz. Unter einer Isospintransformation gehen die Mitglieder eines Isospinmultipletts, etwa des Dubletts (p, n), ineinander über. Eine Invarianz gegenüber dieser Transformation hätte die Konsequenz, daß die Massen innerhalb eines Multipletts übereinstimmen. Die Größe der Massendifferenzen ist ein Maß für die Stärke der Symmetriebrechung.

5.1.3.2. Starke Wechselwirkung. Die starke Wechselwirkung beherrscht die Streu- und Erzeugungsprozesse von Hadronen. Auch die großen anormalen Beiträge zum magnetischen Moment der Baryonen sind ein Effekt der starken Wechselwirkung (Abb. 5.1.3–2). Man spricht in diesem Zusammenhang von einer Wolke virtueller Mesonen rund um das Baryon. Die Stärke der Wechselwirkung drückt sich in Meson-Baryon-Kopplungskonstanten der Größenordnung 10 aus. Es sind jedoch nicht diese Kopplungen zwischen Mesonen und Baryonen, sondern diejenigen der Quantenchromodynamik (vgl. 5.1.2.3.), die als fundamental anzusehen sind. (Abb. 5.1.3–3). Obwohl die Gluonen genau wie das Photon masselos sind, unterscheidet sich die Quantenchromodynamik, die Feldtheorie der starken Wechselwirkungen, wesentlich von der Quantenelektrodynamik. Das gilt insbesondere für die Reichweite der Wechselwirkung (vgl. Tab. 5.1.3–1). Die starken Wechselwirkungen zwischen den Hadronen sind kurzreichweitig, weil die Hadronen im Gegensatz zu den Quarks keine Colorladung tragen. Eine entfernt analoge Situation findet man bei den zwischen den Molekülen wirksamen Van-der-Waals-Kräften, die eine kurze Reichweite haben, weil die Moleküle elektrisch neutral sind.

Von den Problemen, die die Quantenchromodynamik gelöst hat oder noch zu lösen hat, seien hier nur folgende erwähnt. Die experimentellen Ergebnisse der tief-inelastischen Streuung von Leptonen und Nukleonen (vgl. 5.1.4.2.), die Aufschluß über die Struktur der Hadronen erbracht haben, wurden von der Quantenchromodynamik quantitativ erklärt. Eine entscheidende Rolle spielt hierbei die „asymptotische Freiheit" der Quantenchromodynamik: Die Stärke der Wechselwirkung zwischen Quarks verringert sich in dem Maße, wie sich die Quarks einander nähern. Die schwierigere Aufgabe, die Massen der Hadronen u. a. ihrer Eigenschaften (etwa die Verteilung der elektrischen Ladung und des magnetischen Moments) zu berechnen, wurde bisher nur zum Teil und auch nur approximativ gelöst. Weil die Methode der Störungstheorie nicht anwendbar ist, kann man nicht erwarten, daß die Quantenchromodynamik in der Lage sein wird, quantitative Vorhersagen über alle Streu- und Erzeugungsprozesse von Hadronen zu machen. Dessenungeachtet haben die angeführten Teilerfolge ausreichende Belege dafür geliefert, die Quantenchromodynamik als diejenige Feldtheorie anzusehen, die die wesentlichen Aspekte der Hadronen und der starken Wechselwirkung wiederspiegelt.

5.1.3.3. Elektromagnetische Wechselwirkung kann im Falle der Leptonen auf den fundamentalen Elektron-Photon-Vertex und einen analogen für andere Leptonen zurückgeführt werden. Anders bei den Hadronen, wo die elektromagnetische stets mit der starken Wechselwirkung zusammenwirkt. Ein einfaches Beispiel bietet der Proton-Photon-Vertex (vgl. Abb. 5.1.3–2). Zerfallsprozesse, die unter Beteiligung der elektromagnetischen Wechselwirkung ablaufen, zeichnen sich durch Lebensdauern von der Größenordnung 10^{-16} s aus. Beispiele sind $\Sigma^0 \rightarrow \Lambda + \gamma$ und $\Pi^0 \rightarrow 2\gamma$. Wenn dem Photon die Hyperladung $Y = 0$ zugeordnet wird, kann man von einem Erhaltungssatz für Y reden. Beide Prozesse sind auch dadurch bemerkenswert, daß Photonen mit elektrisch neutralen Teilchen wechselwirken. Auf der Grundlage des Quarkmodells kann man das sich so erklären, daß es eigentlich die virtuellen Quarks sind, an die die Photonen koppeln.

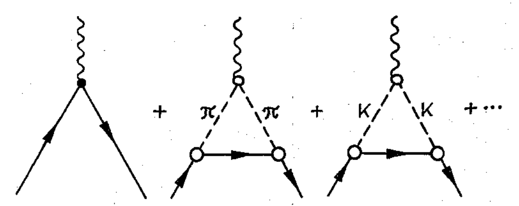


Abb. 5.1.3–2 Beiträge zum Proton-Photon-Vertex; im Falle des Neutrons fehlt der erste Graph

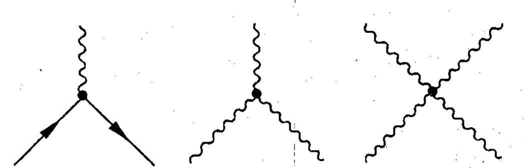


Abb. 5.1.3–3 Die Kopplungen der Quantenchromodynamik. Die Wellenlinien bezeichnen Gluonen

Ein Prozeß, bei dem Y erhalten bleibt, I jedoch nicht, wird im erweiterten Sinne eine elektromagnetische Wechselwirkung genannt, auch wenn kein reelles Photon daran beteiligt ist. Ein Beispiel bietet der Zerfall des η^0-Mesons in drei Pionen. Ein elektromagnetischer Effekt ist auch die Massenaufspaltung innerhalb der Isospinmultipletts. Sie bleibt im Baryonoktett geringer als 0,01 GeV/c^2, das ist eine Größenordnung kleiner als der Abstand der Multipletts untereinander.

5.1.3.4. Schwache Wechselwirkung. Nach der Entdeckung der intermediären Bosonen $W^\pm$, Z^0 steht zweifelsfrei fest, daß die schwache Wechselwirkung auf den Austausch dieser Teilchen zurückzuführen ist. Schwache Wechselwirkungen zwischen Leptonen, etwa der Myon-Zerfall $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$ (Abb. 5.1.3–4a) oder der Streuprozeß $\bar{\nu}_\mu + e^+ \rightarrow \mu^+ + \bar{\nu}_e$

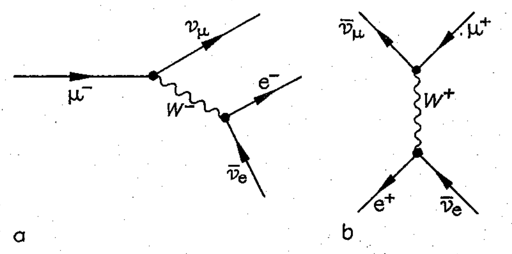


Abb. 5.1.3–4 (a) Hauptbeitrag zum Myonzerfall in der einheitlichen Theorie und (b) analog für den Prozeß $\bar{\nu}\mu + e^+ \rightarrow \mu^+ + \bar{\nu}_e$

(Abb. 5.1.3–4b) werden durch den Austausch eines W-Bosons hinreichend gut beschrieben, obwohl auch hier wie im Falle der Quantenelektrodynamik Strahlungskorrekturen hinzukommen. Wegen der großen Masse von W ist der Beitrag des Diagramms in Abb. 5.1.3–4a zum Myonzerfall quantitativ kaum vom Beitrag einer 4-Fermionen-Kopplung (Abb. 5.1.3–5) zu unterscheiden, und tatsächlich wurde letztere noch unlängst als die fundamentale Kopplung angesehen. Die Stärke dieser approximativ äquivalenten 4-Fermionen-Kopplung wird durch die Kopplungskonstante $G \approx 1{,}1 \cdot 10^{-5}/M_p^2$ angegeben. Die schwache Wechselwirkung zwischen Hadronen führt zu weit komplizierteren Erscheinungen, insbesondere bei den Zerfällen. Charakteristisch ist die große Zahl von Zerfallskanälen (Abb. 5.1.3–6). Die Verzweigungsverhältnisse geben die Wahrscheinlichkeit an, daß der Zerfall in dem betreffenden Kanal stattfindet. Man hat zwischen leptonischen und nichtleptonischen Zerfällen zu unterscheiden. Leptonische Zerfälle gehen unter Aussendung von Leptonen einher, z. B. der β-Zerfall $n \rightarrow p + e^- + \nu_e^-$ und der Zerfall des Pions $\Pi \rightarrow \mu^+ + \nu_\mu$. Beispiele für nichtleptonische Zerfälle sind $\Lambda \rightarrow p + \Pi^-$ oder $\Sigma^- \rightarrow n + \Pi^-$. Die Größe der Lebensdauer von $\approx 10^{-10}$ s zeugt davon, daß auch diese Prozesse durch die schwache Wechselwirkung bedingt sind, obwohl keine Neutrinos ausgesendet werden. In der Regel ändert sich bei nichtleptonischen Zerfällen die Hyperladung um $\Delta Y = 1$, und der Isospin um $\Delta I = 1/2$. Die Graphen dieser Zerfallsprozesse wird man sicher nicht als fundamentale Vertices ansehen können. Fundamentale Kopplungen setzt man statt dessen zwischen Quarks, Leptonen und intermediären Bosonen an. Auf diese Kopplungen kann auch der Prozeß des β-Zerfalls zurückgeführt werden (Abb. 5.1.3–7). Bei nichtleptonischen Prozessen sind in jedem Fall intermediäre Bosonen als virtuelle Teilchen beteiligt.

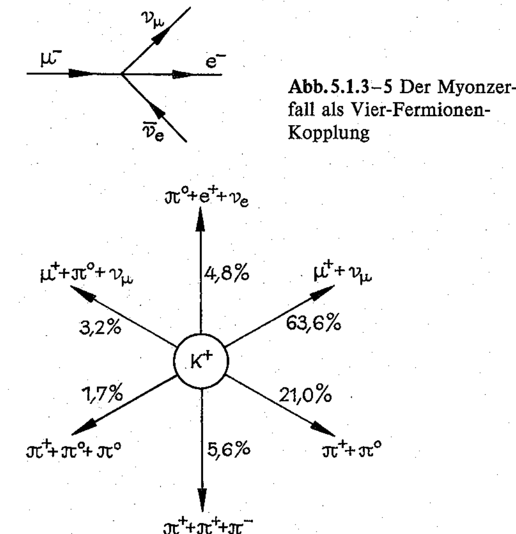


Abb. 5.1.3–5 Der Myonzerfall als Vier-Fermionen-Kopplung

Abb. 5.1.3–6 Zerfallsmöglichkeiten eines positiven K-Mesons mit Angabe der Zerfallswahrscheinlichkeit in die einzelnen Kanäle

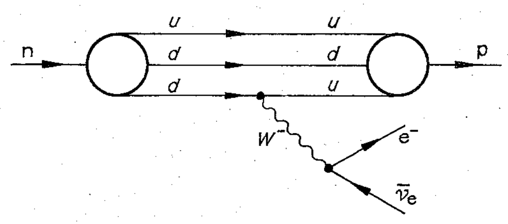


Abb. 5.1.3–7 Ein Beitrag zum β-Zerfall des Neutrons

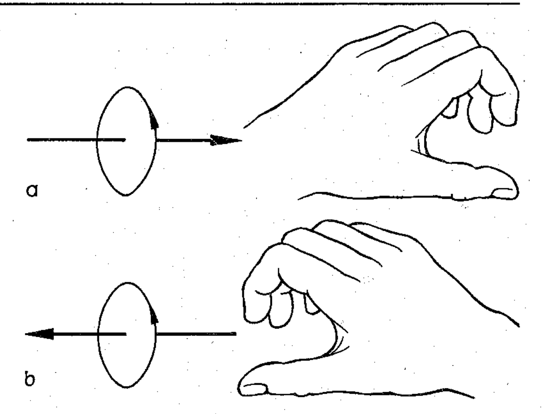


Abb. 5.1.3–8 Orientierung von Spin und Impuls (Pfeilrichtung), (a) bei linkshändigen bzw. (b) rechtshändigen Teilchen

Die Besonderheit der schwachen Wechselwirkung besteht darin, daß die Parität nicht erhalten bleibt (vgl. 5.1.3.1.). Die Struktur dieser Paritätsverletzung zeigt sich besonders klar in denjenigen Reaktionen, die durch den Austausch geladener intermediärer Bosonen beschrieben werden, z. B. $\mu^- \rightarrow e^- + \bar{\nu}_e + \nu_\mu$: Die Neutrinos werden im linkshändigen, die Antineutrinos jedoch im rechtshändigen Zustand erzeugt (Abb. 5.1.3–8). Dasselbe gilt für Reaktionen, an denen Quarks beteiligt sind. Solange die Neutrinos als masselose Teilchen angesehen wurden, ergab sich daraus, daß jedes Neutrino oder Antineutrino links- bzw. rechtshändig ist. Dagegen verliert die Angabe der Händigkeit für massive Teilchen ihren physikalischen Sinn, da jedes massive Teilchen immer nur durch eine quantenmechanische Überlagerung von rechts- und linkshändigen Zuständen zu beschreiben ist. Nach Verlauf einer beliebig kurzen Zeitspanne ist ein im linkshändigen Zustand erzeugtes Neutrino mit einer bestimmten Wahrscheinlichkeit im rechtshändigen Zustand anzutreffen.

Beim Zerfall der neutralen Kaonen wird die *CP*-Invarianz verletzt. Wäre das nicht so, dann dürfte der langlebige Zustand K_L^0 wohl in drei Pionen übergehen, nicht aber in zwei. Das experimentelle Ergebnis ist, daß 0,3 % aller Zustände K_L^0 in ein Paar geladener Pionen zerfällt. Auch die äußerst kleine Differenz zwischen den Massen K_L^0 und K_S^0 in Höhe von etwa $3 \cdot 10^{-12}\,\mathrm{MeV}/c^2$ geht auf die *CP*-verletzende Wechselwirkung zurück, woraus folgt, daß diese offenbar sehr schwach ist. Dieser Effekt läßt sich auf eine Besonderheit des Quarkmodells zurückführen (auf eine besondere Struktur der Quark-Massenterme in den Feldgleichungen). Damit wird allerdings nicht erklärt, warum eine *CP*-Verletzung auftritt. Übrigens erwartet man ähnliches im $D^0\bar{D}^0$-System und im $B^0\bar{B}^0$-System. In den letzten Jahren hat noch eine andere Entdeckung von sich reden gemacht, die unter

dem Schlagwort „neutrale Ströme“ bekannt geworden ist. Es handelt sich hierbei um eine ganz neue Klasse von Prozessen, die die Einführung von zusätzlichen fundamentalen Vertices erforderlich macht. Es wurden Reaktionen beobachtet, die durch Neutrinos ν_μ oder Antineutrinos $\bar{\nu}_\mu$ ausgelöst werden, ohne daß im Endzustand ein Myon oder Antimyon zu finden war. Die normalen Reaktionen lauten dagegen $\nu_\mu + n \to p + \mu^-$ bzw. $\bar{\nu}_\mu + p \to n + \mu^+$. Diese Experimente bewiesen, daß es auch die Reaktionen $\nu_\mu + n \to \nu_\mu + n$ und $\bar{\nu}_\mu + p \to \bar{\nu}_\mu + p$ gibt. Der Umstand, daß das Lepton im Gegensatz zu den normalen Reaktionen seine elektrische Ladung nicht ändert, führte zu der Bezeichnung „neutrale Ströme“. Besonders eindrucksvoll wurde die Existenz zusätzlicher Kopplungen durch die Beobachtung des Prozesses $\bar{\nu}_\mu + e^- \to \bar{\nu}_\mu + e^-$ belegt. Nach landläufigen Vorstellungen wäre dieser Prozeß auszuschließen, da Antineutrinos $\bar{\nu}_\mu$ nur mit Positronen reagieren (vgl. Abb. 5.1.3–4b). Diese Prozesse werden durch den Austausch der neutralen intermediären Bosonen Z^0 beschrieben. Die Entdeckung der neuen Prozesse ist darum so bedeutsam, weil ihre Existenz eine unerläßliche Voraussetzung für das Programm einer einheitlichen theoretischen Behandlung der elektromagnetischen und schwachen Wechselwirkungen ist.

5.1.3.5. Gravitative Wechselwirkung. Ohne Zweifel unterliegen auch die Elementarteilchen der Gravitation, d. h. der universellen gegenseitigen Anziehung aller Massen. Der Effekt ist allerdings äußerst klein, so daß man bei gegenwärtigen Experimenten im Laboratorium völlig von der Gravitation absehen kann. Die Schwäche dieser Wechselwirkung wird ersichtlich, wenn man die zwischen zwei Protonen wirkende elektrische Coloumb-Kraft mit der Massenanziehung vergleicht. Es ergibt sich dafür das Verhältnis $(e^2/4\pi\varepsilon_0)/\kappa M_p^2 \sim 10^{36}$. Für die Gravitationskonstante κ gilt $\kappa \approx 6{,}67 \cdot 10^{-11} \mathrm{m^3 kg^{-1} s^{-2}}$. Bei sehr hohen Energien wird jedoch die Gravitation nicht immer zu vernachlässigen sein. Das liegt anschaulich ausgedrückt daran, daß auch die Energie entsprechend der Einsteinschen Beziehung $E = mc^2$ gravitierend wirkt und somit eine hinreichende Vergrößerung der Teilchenenergie schließlich die Kleinheit der Gravitationskonstante zu kompensieren vermag. Auch Probleme der kosmischen Physik, z. B. die Schwarzen Löcher und ihr Zerfall, weisen auf eine enge Verflechtung von Elementarteilchen und Gravitation hin. Eine konsistente theoretische Behandlung beider Gebiete, die eine Quantentheorie der Gravitation einschließen muß, ist eine Aufgabe von grundlegender Bedeutung.

5.1.3.6. Einheitliche Theorie der elektromagnetischen und schwachen Wechselwirkungen. Die einheitliche Theorie der elektromagnetischen und schwachen Wechselwirkungen kann nach der Entdeckung der intermediären Bosonen W^+ und Z^0 als bestätigte Theorie angesehen werden, wodurch insbesondere die früheren Vorstellungen über den Mechanismus der schwachen Wechselwirkung grundlegend geändert worden sind. Daß die schwache Wechselwirkung durch den Austausch intermediärer Bosonen vermittelt wird, stellt sie der elektromagnetischen Wechselwirkung an die Seite. Komplikationen werden durch die von Null verschiedenen Massen von W^+ und Z^0 sowie die paritätsverletzende Struktur der schwachen Wechselwirkung hervorgerufen.

Diese Links-Rechts-Asymmetrie macht es erforderlich, den links- und rechtshändigen Komponenten der Lepton- und Quarkfelder (vgl. 5.1.2.5.) unterschiedliche innere Quantenzahlen zuzuordnen. Die hier wesentlichen Quantenzahlen tragen die Bezeichnungen „schwacher Isospin“ I^W und „schwache Hyperladung“ Y^W; beide haben erst im Rahmen der einheitlichen Theorie Bedeutung erlangt. Der schwache Isospin gliedert die Fermionenfelder in Singletts, Dubletts, Tripletts ähnlich, wie es der eigentliche Isospin mit Hadronen und Quarks tut. Die linkshändigen Komponenten der Leptonen (e_L^-, ν_{eL}, μ_L^-, $\nu_{\mu L}$, τ_L^-, $\nu_{\tau L}$) und der Quarks (u_L, d_L, c_L, s_L, t_L, b_L) ordnen sich in Dubletts des schwachen Isospins, die (analog bezeichneten) rechtshändigen Komponenten in Singletts (Tab. 5.1.3–9). Zwischen der elektrischen Ladung Q, der dritten Komponente des schwachen Isospins I_3^W und der schwachen Hyperladung Y^W besteht die Relation $Q = I_3^W + 1/2\, Y^W$.

Tab. 5.1.3–9 Quantenzahlen der fundamentalen Fermionen in der einheitlichen Theorie

	ν_{eL} $\nu_{\mu L}$ $\nu_{\tau L}$	e_L μ_L τ_L	e_R μ_R τ_R	u_L c_L t_L	d_L s_L b_L	u_R c_R t_R	d_R s_R b_R
I_3^W	+1/2	−1/2	0	+1/2	−1/2	0	0
Y^W	−1	−1	−2	1/3	1/3	4/3	−2/3

Die einheitliche Theorie beschreibt die Wechselwirkung der Fermionenfelder durch die Kopplung an vier Vektorfelder (d. h. Felder, deren Teilchen den Spin 1 tragen). Die linkshändigen Dubletts (ν_{eL}, e_L^-), ($\nu_{\mu L}$, μ_L^-), ($\nu_{\tau L}$, τ_L^-), ... koppeln an das I^W-Triplett der Vektorfelder (W^+, W^0, W^-) und an das Singlett-Vektorfeld B^0, die rechtshändigen Singletts e_R^-, u_R, μ_R^-, ... dagegen nur an das Singlett-Vektorfeld B^0. Die Stärke der Kopplung an B^0 ist proportional zum Wert der Hyperladung der Fermionen. Die Kopplungen an das Triplett und an das Singlett erfolgen mit zwei unabhängigen Kopplungskonstanten.

Die eigentliche Komplikation liegt nun darin, daß die neutralen Felder W^0 und B^0 noch keine physikalischen Felder sind, denen Teilchen im üblichen Sinne entsprechen. Vielmehr muß man aus ihnen die Kombinationen $A = B^0 \cos\Theta_W + W^0 \sin\Theta_W$ und $Z^0 = -B^0 \sin\Theta_W + W^0 \cos\Theta_W$ bilden, um zu den Feldern für das Photon und das Z^0 zu gelangen. In diesen Beziehungen drückt sich eine enge Verwandtschaft zwischen Photon und Z^0 aus, was einer der Gründe dafür ist, die vorliegende Theorie als eine einheitliche Theorie beider Wechselwirkungen zu bezeichnen.

Der Winkel Θ_W heißt *Weinberg-Winkel*, er ist mit den Massen durch die Relation $M_Z = M_W / \cos\Theta_W$ verknüpft. Der Weinberg-Winkel ist schon vor dem Nachweis der intermediären Bosonen in verschiedenen, voneinander unabhängigen Experimenten gemessen worden. Die eine Gruppe von Experimenten zielt auf den Vergleich der Wirkungsquerschnitte von Prozessen ab, die entweder durch geladene oder durch neutrale Ströme vermittelt werden. Bei der anderen Gruppe handelt es sich um die Bestimmung des relativen Anteils von Z^0-Austausch in Prozessen, die vorwiegend über Photonenaustausch ablaufen, z. B. bei der tief-inelastischen Streuung von Elektronen oder Myonen an Nukleonen (vgl. 5.1.4.2.). Die übereinstimmenden Ergebnisse dieser Experimente weisen nachdrücklich darauf hin, daß die durch den Weinberg-Winkel definierte Abweichung der physikalischen von den ursprünglichen Feldern in der Natur der Dinge liegt, so willkürlich und künstlich sie auch erscheinen mag.

Die Unterschiede in den Massen von $W^\pm$, Z^0 und γ sind ein

Ausdruck der Brechung der mit dem schwachen Isospin (und der schwachen Hyperladung) verknüpften Symmetrie. Die unmittelbare Einfügung von Massengliedern in die Feldgleichungen von Vektorfeldern hätte die Renormierbarkeit (vgl. 5.1.1.9.) zerstört. Einen Ausweg bietet erst der Mechanismus der „spontanen Symmetriebrechung". Zu diesem Zweck wird neben den Fermionen und den zunächst masselosen intermediären Bosonen zusätzlich ein Dublett (Φ_1, Φ_2) mit dem Spin 0 (das sogenannte *Higgs-Feld*) eingeführt, dessen Wechselwirkung durch ein speziell gewähltes Potential $V(\Phi)$ (Abb. 5.1.3–10) bestimmt ist. Das Minimum des Potentials gibt den Zustand niedrigster Energie, d.h. den Grundzustand an; dieser ist im vorliegenden Fall durch von Null verschiedene Felder Φ_1 oder Φ_2' charakterisiert, es liegt ein sogenanntes Kondensat vor. Dieses Kondensat führt zu Massengliedern in den Feldgleichungen der Vektorfelder, ohne daß die Renormierbarkeit beeinträchtigt wird.

Die einheitliche Theorie ist wie die Quantenchromodynamik eine Eichtheorie, für die insbesondere die Kopplungen der intermediären Bosonen untereinander typisch sind. Zu den in Abb. 5.1.3–11 angegebenen Kopplungen kommen noch alle diejenigen hinzu, in denen das Fermionpaar (ν_e, e^-) durch andere Paare (ν_μ, μ), (u, d) usw. ersetzt ist. Außerdem sind die Kopplungen an das Higgs-Feld hinzuzufügen. Der gründliche experimentelle Test der einheitlichen Theorie wird an den e^+e^--Speicherringen (vgl. 5.1.4.1.) der neuen Generation erfolgen. Hier wird auch nach dem physikalischen Higgs-Teilchen gesucht werden, einem elektrisch neutralen Teilchen mit dem Spin 0, dessen Masse die Theorie nicht vorhersagen kann.

5.1.3.7. Große Unifikation – Superschwache Wechselwirkung. Die große Unifikation verfolgt das Ziel, sowohl die starken als auch die elektromagnetischen und schwachen Wechselwirkungen aus einem gemeinsamen Ursprung zu begreifen. Sie folgt dem Vorbild der einheitlichen Theorie der elektromagnetischen und schwachen Wechselwirkungen (vgl. 5.1.3.6.) insbesondere darin, daß die Idee der spontanen Symmetriebrechung zugrunde gelegt wird. Das wesentliche Kennzeichen der großen Unifikation ist die Zusammenfassung der fundamentalen Fermionenfelder (zumindest einer jeden Familie) in große Multipletts, die an einen erweiterten Satz intermediärer Bosonen koppeln, wodurch in jedem Falle Umwandlungen von Quarks in Leptonen ins Spiel kommen. Die Theorie läßt somit Prozesse zu, bei denen die Baryonzahl nicht erhalten bleibt. Da andererseits das Proton im praktischen Sinne stabil ist (seine Lebensdauer beträgt mehr als 10^{31} Jahre, vgl. 5.1.4.3.), ergeben sich sofort starke Einschränkungen für die Theorie. Es muß vorangestellt werden, daß sich sehr verschiedene „große einheitliche Theorien" konstruieren lassen, die sich sowohl hinsichtlich der Struktur der Fermionenmultipletts als auch der Zahl der intermediären Bosonen und der die Symmetriebrechung bewerkstelligenden Higgs-Felder unterscheiden.

Die einfachste denkbare Variante ordnet die Fermionen einer Familie (genauer, ihre Rechts- und Linkskomponenten) in ein Fünfer- und ein Zehnermultiplett. Die Wechselwirkungen werden durch 24 intermediäre Bosonen vermittelt; das sind neben den schon bekannten (γ, W^+, W^-, Z^0, acht Gluonen) die neuen intermediären Bosonen X und Y, die die elektrische Ladung 4/3 bzw. 1/3 tragen. Ein zum Protonzerfall beitragender Feynman-Graph ist in Abb. 5.1.3–12 gezeigt. Die Übergangswahrscheinlichkeit muß in Einklang mit der Protonlebensdauer, $\tau_p > 10^{31}$ Jahre, sehr gering sein, woraus sich unmittelbar die Abschätzung $M_X \sim M_Y \gtrsim 10^{15}\,\text{GeV}/c^2$ ergibt. Die durch den Austausch von X oder Y erfolgenden Prozesse können also mit Recht als superschwache Wechselwirkungen bezeichnet werden.

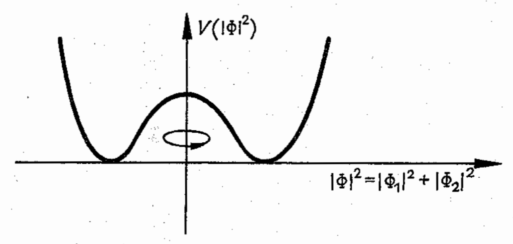


Abb. 5.1.3–10 Die potentielle Energie des Higgs-Feldes

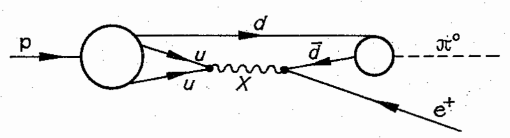


Abb. 5.1.3–11 Fundamentale Kopplungen der einheitlichen Theorie

Eine zweite unabhängige Abschätzung der Massen der schweren intermediären Bosonen folgt aus der Untersuchung des Zusammenhangs, der zwischen den drei Kopplungskonstanten der QCD und der einheitlichen Theorie der elektromagnetischen und schwachen Wechselwirkungen einerseits, und der einen Kopplungskonstante der großen einheitlichen Theorie andererseits besteht. Wäre die Symmetrie nicht gebrochen, so beständen zwischen den drei Kopplungskonstanten eindeutig bestimmte Relationen; auch der Weinberg-Winkel (vgl. 5.1.3.6.) wäre aus der großen Unifikation unmittelbar zu berechnen. Tatsächlich aber ist die Symmetrie gebrochen; jedoch wird die Symmetriebrechung keine Rolle spielen bei solchen Prozessen, bei denen die Energien aller beteiligten Teilchen größer als M_X und M_Y sind. Der Formalismus der Quantenfeldtheorie läßt eine Umrechnung aller zwischen Kopplungskonstanten bestehenden Relationen und auch des Weinberg-Winkels auf typische Laboratoriumsenergien zu. Damit ergibt sich unter Berücksichtigung der Symmetriebrechung ein Wert für den Weinberg-Winkel, der mit dem experimentell bestimmten Wert übereinstimmt, falls $M_X \sim M_Y \sim 10^{15}\,\text{GeV}/c^2$ angenommen wird.

Es ist vollkommen aussichtslos, die Verifikation der großen

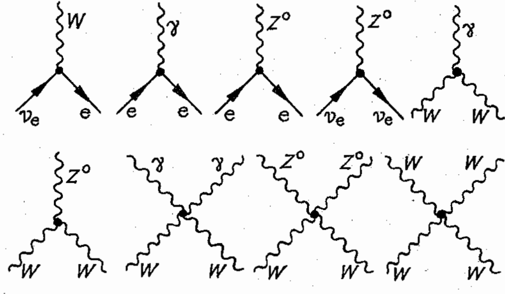


Abb. 5.1.3–12 Einer der zum Protonzerfall $p \rightarrow e^+ + \pi^0$ beitragenden Graphen

Unifikation an den Nachweis der schweren intermediären Bosonen knüpfen zu wollen. Diese können im Laboratorium nicht erzeugt werden, da auch die in der weiteren Perspektive verfügbaren Beschleunigerenergien dazu nicht ausreichen werden. Außerdem wären solche Teilchen extrem kurzlebig, so daß sie ebensowenig als Relikte aus der Frühphase des Kosmos (vgl. 5.1.3.8.) übriggeblieben sein könnten. Nur der Nachweis des Protonzerfalls kann darüber entscheiden, ob die Idee der großen Unifikation in der Natur realisiert ist. Die genaue Bestimmung der Lebensdauer und der Verzweigungsverhältnisse des Zerfalls würde dann auch einen Rückschluß darauf zulassen, welche der denkbaren großen einheitlichen Theorien vorliegt.

5.1.3.8. Verbindungen zur Kosmologie. Im Rahmen der Vorstellungen, die den Beginn der Entwicklung des Kosmos in einem Zustand hoher Temperatur und hoher Dichte sehen, ergibt sich zwangsläufig, daß die Frühphase des Kosmos wesentlich durch die Elementarteilchenphysik bestimmt worden ist. Ebenso wie feste Körper, Moleküle und Atome bei hinreichend hoher Temperatur in ihre Bestandteile dissoziieren, erwartet man eine Dissoziation der Hadronen in freie Quarks. Theoretische Untersuchungen der Quantenchromodynamik haben Hinweise auf eine solche Dissoziation erbracht, die bei Temperaturen oberhalb einiger 100 MeV erfolgen sollte. Es ist hier gebräuchlich, die Temperaturen gemäß $E = kT$ (k = Boltzman-Konstante) in Einheiten der Energie auszudrücken: 100 MeV $\triangleq 1{,}2 \cdot 10^{12}$ K. Solange also die Temperatur in dieser Größenordnung lag, existierte die Materie in Form eines Plasmas aus Quarks, Leptonen und intermediären Bosonen.

Es besteht übrigens ein wesentlicher Unterschied zwischen der Dissoziation eines Atomkerns in Nukleonen und der Dissoziation eines Hadrons in Quarks. Während ein Atomkern in hinreichend energiereichen Stoßprozessen in seine Bestandteile (Nukleonen) zertrümmert werden kann, ist das mit Hadronen nicht möglich. Dazu müssen erst die Bedingungen eines thermischen Gleichgewichtszustandes vorliegen, der durch hinreichend hohe Teilchenzahl und eine Gleichverteilung der Energie $kT/2$ auf alle Freiheitsgrade charakterisiert ist. Ein solcher Zustand ist in typischen Streuexperimenten nicht realisiert; man hofft jedoch, ihn in den Stößen schwerer Ionen zu erreichen.

Ein weiterer wichtiger Temperatureffekt betrifft die spontane Symmetriebrechung (vgl. 5.1.3.6.). Quantenfeldtheoretische Korrekturen verändern bei hinreichend hoher Temperatur das Higgs-Potential (vgl. Abb. 5.1.3–10) derart, daß das Minimum in den Nullpunkt rutscht und die Vorbedingung der spontanen Symmetriebrechung somit aufgehoben wird (Abb. 5.1.3–13). Wenn also eine große einheitliche Theorie als gültig angenommen wird, so kann daraus geschlossen werden, daß bei anfänglichen Temperaturen von über 10^{15} GeV $\triangleq 10^{28}$ K eine ungebrochene Symmetrie zwischen Quarks und Leptonen und auch aller intermediären Bosonen untereinander bestanden hat.

Die große Unifikation macht jedoch nicht nur Aussagen über einen unvorstellbar weitentrückten Anfangszustand des Kosmos. Sie ist vielmehr auch in der Lage, zum Verständnis der Eigenschaften des Kosmos beizutragen, die sich uns durch gegenwärtige Beobachtungen erschließen: Als ein Beispiel sei die makroskopische Baryon-Antibaryon-Asymmetrie genannt. Darunter versteht man das Fehlen irgendwelcher Anzeichen für die Existenz von Antimaterie in makroskopischen Ausmaßen. Man kann jedoch nicht erwarten, aus den beobachtbaren Fakten der Kosmologie Rückschlüsse auf die Gültigkeit dieser oder jener Variante der großen einheitlichen Theorie ziehen zu können; das bleibt allein den Experimenten in irdischen Labors vorbehalten.

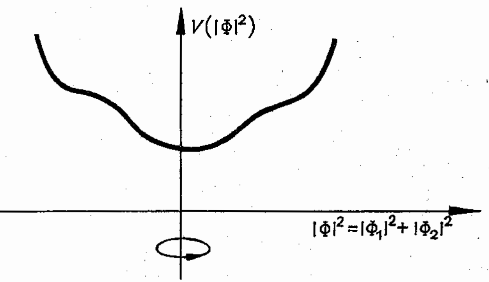


Abb. 5.1.3–13 Die potentielle Energie des Higgs-Feldes bei Temperaturen oberhalb der kritischen

Weitere Querverbindungen zwischen Elementarteilchenphysik und Kosmologie sind mit dem Problem der Neutrinomassen verknüpft. Die elektromagnetische kosmische Untergrundstrahlung, die sogenannte 3-Kelvin-Strahlung, ist nicht das einzige Relikt aus der Frühphase des ehemals heißen Kosmos. Es muß ebenso eine im thermischen Gleichgewicht befindliche Neutrino-Antineutrino-Strahlung existieren, deren Teilchendichte in einem bestimmten Verhältnis zur Dichte der Photonen der 3-Kelvin-Strahlung steht. Aus der experimentell bestimmten Photondichte $N_\gamma \sim 500/\text{cm}^3$ schließt man auf eine Zahl von ≈ 400 Neutrinos und Antineutrinos aller drei Familien je Kubikzentimeter. Wenn allein für das Elektron-Neutrino eine Masse von 20 eV/c^2 angenommen wird, kann gefolgert werden, daß die Neutrinos den überwiegenden Beitrag zur mittleren Massendichte des Kosmos liefern. Darüber hinaus könnten die Beiträge der Myon-Neutrinos und der Tau-Neutrinos die mittlere Massendichte sehr leicht in die Nähe desjenigen Grenzwertes bringen, der die zwei Fälle der unbeschränkten Expansion des Kosmos und des oszillierenden Kosmos voneinander trennt. So ergibt sich der erstaunliche Sachverhalt, daß Ergebnisse von Laboratoriumsexperimenten an Elementarteilchen eine entscheidende Rolle bei den Vorhersagen über das Schicksal des Kosmos in ferner Zukunft spielen werden.

5.1.4. Wichtige Prozesse der Elementarteilchenphysik

In diesem Kapitel werden drei Prozesse vorgestellt, deren Untersuchung richtungsweisende Ergebnisse erwarten lassen. Der Prozeß der Elektron-Positron-Paarvernichtung wird in absehbarer Zukunft die wichtigste Quelle zur Erzeugung und damit auch zur Entdeckung neuer Teilchen sein. Die Untersuchung der tief-inelastischen Lepton-Hadron-Streuung wird die Erforschung der Nukleonstruktur in noch kleinere Raumbereiche fortsetzen. Die Experimente zum Protonzerfall sollen endlich ein abschließendes Urteil über die Ideen der Großen Unifikation ermöglichen.

5.1.4.1. Elektron-Positron-Vernichtung. Dieser Prozeß ist von besonderer Bedeutung, da die Mehrzahl neuer Teilchen auf diesem Wege entdeckt wurde, und auch in Zukunft überraschende Entwicklungen nicht auszuschließen sind. Außerdem eröffnet die Elektron-Positron-Vernichtung die Möglichkeit, Aufschluß über die Zahl der Quarks zu erhalten. Der Prozeß der Paarvernichtung wird an den Elek-

tron-Positron-Speicherringen untersucht (Abb. 5.1.4–1). Der dynamische Mechanismus der Paarvernichtung ist dadurch gekennzeichnet, daß das Leptonpaar nur durch virtuelle Photonen mit den Teilchen des Endzustands in Wechselwirkung stehen kann. Wegen der Kleinheit der elektromagnetischen Kopplungskonstanten ist es der Austausch eines einzigen Photons, der den Prozeß in guter Näherung beschreibt. Die Erzeugung von Hadronen stellt man sich im Quarkmodell so vor, daß das virtuelle Photon an ein virtuelles Quark-Antiquark-Paar koppelt, das schließlich in einen reellen Hadronzustand übergeht (Abb. 5.1.4–2a). Bei höheren Energien, d.h. für Energien größer als 2×20 GeV, wird gleichermaßen der Austausch des neutralen interme-

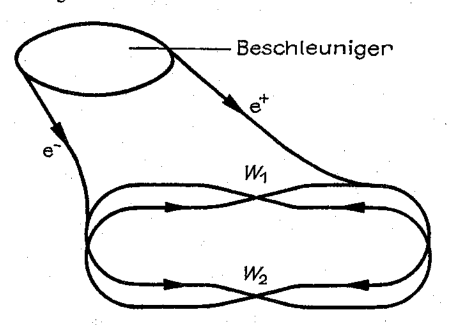


Abb. 5.1.4–1 Schema von Speicherringen; aus dem Beschleuniger werden Elektronen und Positronen in die vertikal versetzten Speicherringe eingespeist; die Strahlen treffen in W_1 und W_2 aufeinander

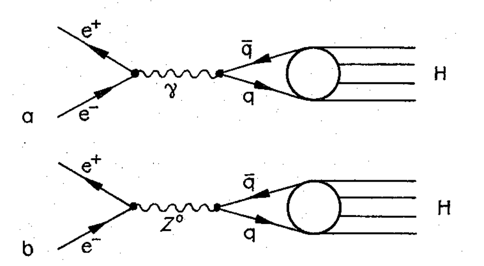


Abb. 5.1.4–2 Erzeugung von Hadronen H durch Elektron-Positron-Vernichtung

diären Bosons Z^0 wichtig (Abb. 5.1.4–2b). Solange nur der Ein-Photon-Austausch zu berücksichtigen ist, ergibt sich eine einfache theoretische Aussage über den Quotienten aus dem Wirkungsquerschnitt der Hadronerzeugung und dem Wirkungsquerschnitt der Reaktion $e^+ + e^- \rightarrow \mu^+ + \mu^-$. Dieser Quotient R sollte für große Energien gegen den Wert $\sum Q_i^2$ streben. Hierbei bedeuten Q_i die elektrischen Ladungen der Quarks, gemessen in Einheiten der Elementarladung. Bei Berücksichtigung der in Tab. 5.1.2–4 aufgeführten Quarks findet man $\sum Q_i^2 = 3\{4/9 + 1/9 + 4/9 + 1/9 + 4/9 + 1/9\} = 5$. Der Faktor 3 berücksichtigt, daß jede Quark in drei verschiedenen Colorzuständen auftreten kann. Dagegen wächst bei einer Energie von ≈ 28 GeV der gemessene Quotient R auf den Wert $R \approx 4$ an. In diesem Energiebereich manifestieren sich demnach nur die fünf Quarks u, d, c, s und b.

In die Elektron-Positron-Speicherringe setzt man weltweit große Erwartungen. Solche Anlagen werden 1986 mit der Energie von 2×30 GeV zur Verfügung stehen (Japan), und ein Jahr später schon mit 2×50 GeV (USA), während sich das bedeutendste Vorhaben im Europäischen Kernforschungszentrum Genf mit der projektierten Energie von 2×100 GeV im Bau befindet. An diesen Speicherringen wird auch eine umfassende Prüfung der einheitlichen Theorie der elektromagnetischen und schwachen Wechselwirkungen möglich sein.

5.1.4.2. Tief-inelastische Streuung von Leptonen. Diese Experimente dienen der eigentlichen Erforschung der Hadronstruktur. Praktisch handelt es sich um die Streuung von Myonen oder Myon-Neutrinos an Nukleonen, wobei sowohl die Energie E der einfallenden Leptonen (100 GeV und mehr), als auch die sogenannte Impulsübertragung auf das Nukleon groß sind. Hierbei ist die Impulsübertragung $\sqrt{Q^2}$ durch die relativistische Formel $Q^2 = (p - p')^2 - (E - E')^2$ bestimmt, wobei p' und E' Impuls bzw. Energie des gestreuten Leptons bezeichnen. Die Umschreibung „tief-inelastisch" drückt zuerst einmal aus, daß der Prozeß ein wesentlich inelastischer ist: das Nukleon geht i. allg. in ein komplexes System von Hadronen über. Ganz anders als bei der Bestimmung der räumlichen Ausdehnung wird das Objekt der Strukturbestimmung zerstört. Die großen Werte der Impulsübertragung ermöglichen den Einblick in die innere Struktur der Hadronen. Auch bei diesem Prozeß kann die elektromagnetische oder schwache Wechselwirkung der Leptonen näherungsweise auf die niedrigste Ordnung der Kopplungskonstanten beschränkt werden. Im Falle der Streuung von Myonen wird der Prozeß demnach durch den Feynman-Graphen (Abb. 5.1.4–3) beschrieben.

Die bisher durchgeführten Experimente zeigten ein überraschendes Ergebnis. Es wurden Wirkungsquerschnitte beobachtet, wie sie am ehesten bei der Streuung von Leptonen an punktförmigen Teilchen zu erwarten wären. Darüber hinaus ist ein Rückschluß auf den Spin dieser Streuzentren möglich; er beträgt $J = 1/2$. Die erste Schlußfolgerung ist, daß der tief-inelastische Streuprozeß nicht am Hadron als Ganzem erfolgt, sondern an irgendwelchen Bestandteilen des Hadrons, die man, vorerst ohne genaue Kenntnis ihrer Natur, als *Partonen* bezeichnet. Eine Streuung am Nukleon als Ganzem liegt bei der elastischen Streuung vor; dort zeigt sich sehr deutlich, daß das Nukleon nicht punktförmig ist. Als zweites liegt die Vermutung sehr nahe, die Partonen mit den Quarks zu identifizieren. Im Prinzip kann auch diese Frage durch das Experiment beantwortet werden. Die spezifischen fundamentalen Kopplungen zwischen Quarks und Photonen bzw. zwischen Quarks und Leptonen führen nämlich zu bestimmten Beziehungen zwischen den Wirkungsquerschnitten der Streuung von Myonen, Neutrinos und Antineutrinos an Nukleonen, den sogenannten Summenregeln. Die bisher vorliegenden experimentellen Ergebnisse lassen innerhalb ihrer Fehlergrenzen den Schluß zu, daß die Partonen nichts anderes als Quarks sind. Es ist auch möglich, die Quarkanteile quantitativ durch Verteilungsfunktionen zu beschreiben. Man stelle sich ein

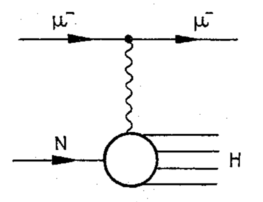


Abb. 5.1.4–3 Tief-in-elastische Myon-Nukleon-Streuung

Nukleon mit dem Impuls p vor. Zu diesem Gesamtimpuls tragen die Quarks mit dem Bruchteil X, $(0 < x < 1)$, bei. Der Anteil der einzelnen Quarksorten u, d usw. kann durch die statistischen Verteilungsfunktionen $U(x)$, $D(x)$ usw. erfaßt werden. Zum Beispiel gibt $U(x)\,\Delta x$ die Wahrscheinlichkeit an, im Nukleon ein u-Quark mit einem zwischen xp und $(x+\Delta x)p$ liegenden Impuls zu finden. Abb. 5.1.4–4 zeigt Verteilungsfunktionen, die mit den experimentellen Resultaten verträglich sind. Ein eindeutiger Schluß auf die einzelnen Kurven ist allerdings nicht möglich. Es gibt Hinweise darauf, daß diese Vorstellungen noch zu naiv sind. Der deutlichste ist die Bilanz für den Nukleonimpuls, die unter Zuhilfenahme der gegebenen Quarkverteilungsfunktionen auf ein Defizit von $\approx 50\,\%$ führt!

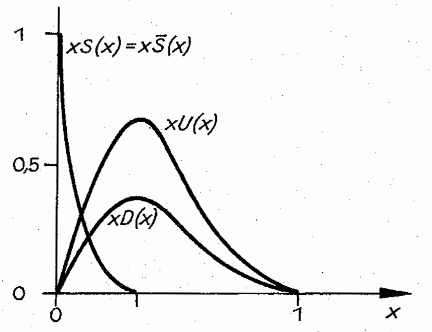


Abb. 5.1.4–4 Gegenwärtige Vorstellungen von den Quarkverteilungsfunktionen des Protons

Eine befriedigende Erklärung für diese u. a. Besonderheiten hat die Quantenchromodynamik (vgl. 5.1.3.2.) erbracht.
Die Meßapparaturen der tief-inelastischen Streuung zeichnen sich durch ungewöhnliche Dimensionen aus. So hat etwa das im Kernforschungszentrum Genf betriebene Myonspektrometer eine Länge von über 50 m und besteht aus einigen Hundert großen Zählerkammern und 1 600 t Magneteisen. Der bisher erforschte Energiebereich ist durch die Parameter $E \leqq 280$ GeV und $Q^2 \leqq 300\,[\mathrm{GeV}]^2/c^2$ charakterisiert. Wenn die in verschiedenen Ländern bestehenden Pläne zum Bau größerer Protonenbeschleuniger verwirklicht werden, dürften zu Beginn der neunziger Jahre tief-inelastische Streuexperimente bei Impulsübertragungen von einigen $1\,000\,[\mathrm{GeV}]^2/c^2$ möglich sein. Zu diesem Zweck müssen allerdings auch die Meßapparaturen verfeinert werden, weil die zu erwartenden Wirkungsquerschnitte mit steigender Impulsübertragung sehr stark abnehmen. Das vorrangige Ziel dieser Experimente wird es sein, die Nukleonstruktur in noch kleineren Raumbereichen zu untersuchen, und nachzuprüfen, ob die Quarks selbst vielleicht noch eine auflösbare Struktur zeigen.

5.1.4.3. Experimente zum Protonzerfall. Der Nachweis des Protonzerfalls könnte das Schlüsselexperiment zur Bestätigung der großen Unifikation (vgl. 5.1.3.7.) sein; deshalb haben solche Experimente in den letzten Jahren eine große Bedeutung erlangt. Es geht darum, einen Zerfall nachzuweisen, dessen Halbwertszeit größer als 10^{31} Jahre ist, und damit mehr als das 10^{21}-fache des Alters des Kosmos beträgt. Die statistische Unabhängigkeit von Quantenprozessen gestattet es jedoch, hierbei Zerfallsraten von einigen Hundert je Jahr zu erwarten, falls nur die unter Beobachtung stehende Materiemenge hinreichend groß ist, d. h., hinreichend viele Protonen enthält. So sind in einer Menge von 1 000 t Wasser 60 Nukleonzerfälle je Jahr zu erwarten, wenn eine Lebensdauer von 10^{31} Jahren angenommen wird. In allen denkbaren Zerfällen des Protons, von denen hier nur die Prozesse $p \to e^+ + \pi^0$, $\mu^+ + \pi^0$, $\bar{\nu} + \pi^+$, $\bar{\nu} + K^+ \ldots$ genannt seien, tritt im Endzustand ein elektrisch geladenes Teilchen mit hoher kinetischer Energie auf, dessen Nachweis als Signal des erfolgten Zerfalles dient. Das Hauptproblem besteht nun darin, derartige Teilchen auszusondern, die nicht von Protonzerfällen herrühren. Jede an der Erdoberfläche befindliche Nachweisapparatur wäre dem Strom der aus den oberen Schichten der Atmosphäre stammenden Myonen ausgesetzt, die die Signale der seltenen Protonzerfälle hoffnungslos überdecken würden. Deshalb werden alle Experimente zum Protonzerfall in Bergwerkstollen oder Gebirgstunneln durchgeführt, wodurch der Myonstrom weitgehend abgeschirmt wird. Um die wenigen übrigbleibenden Myonen als von außen kommende Fremdteilchen identifizieren zu können, sind die Anlagen allseitig von einer Schicht Zählerkammern umgeben (vgl. 5.1.1.2.).
Ein Teil der Myonen erzeugt noch in der Atmosphäre infolge des Zerfallsprozesses $\mu^- \to e^- + \bar{\nu}_e + \nu_\mu$ energiereiche Neutrinos, gegen die keinerlei Abschirmung möglich ist. Diese sogenannten atmosphärischen Neutrinos können in der Meßapparatur infolge der Reaktionen ν_μ + Nukleon $\to \mu^- + \ldots$ und $\bar{\nu}_e$ + Nukleon $\to e^+ + \ldots$ geladene energiereiche Leptonen erzeugen, die den Zerfall von Protonen vortäuschen. Eine Abschätzung läßt erwarten, daß auf ≈ 20 derartige Signale höchstens ein echter Zerfallsprozeß kommt. Man muß deshalb jedes einzelne Ereignis analysieren, um vorgetäuschte Zerfälle mit höchstmöglicher Sicherheit ausschließen zu können.
Die bis 1984 in Betrieb genommenen Meßapparaturen unterteilen sich in zwei Typen. Bei den sogenannten Wasser-Čerenkov-Anlagen ist das Wasser sowohl Beobachtungsobjekt als auch Teil der Nachweisapparatur, werden doch die geladenen energiereichen Teilchen durch ihre im Wasser erzeugte Čerenkov-Strahlung (vgl. 5.1.1.2.) nachgewiesen. Der wichtigste Teil der Apparatur sind Fotovervielfacher (in größeren Anlagen dieser Art über 2 000 Stück), die die schwachen Lichtsignale verstärken und zugleich eine Lokalisierung des Ereignisses ermöglichen. Der andere Typ von Meßapparaturen besteht aus einer größeren Anzahl von ≈ 1 cm dicken Eisenplatten, die die unter Beobachtung stehenden Nukleonen präsentieren, zwischen denen jeweils Zählerkammern (vgl. 5.1.1.2.) angeordnet sind, um die Spuren der Zerfallsprodukte zu registrieren. Die unter Beobachtung stehende Materiemenge beträgt einige 1 000 t Wasser bzw. einige 100 t Eisen.
Das Ziel dieser Experimente ist über den Nachweis des Zerfalls hinaus eine Bestimmung der Verzweigungsverhältnisse, was einen Rückschluß auf die spezielle Art der zugrunde liegenden großen einheitlichen Theorie ermöglichen würde.

5.1.5. Hypothetische Teilchen

Die Weiterentwicklung der Elementarteilchenphysik wird zeitweise durch die Ausbildung neuer theoretischer Modelle bestimmt, die zuerst nur durch ihre interessanten mathematischen Eigenschaften motiviert sind. Da es sich auch hierbei um Quantenfeldtheorien handelt, kommen neue, zunächst hypothetische Teilchen ins Spiel, deren Entdeckung der erste Schritt zur experimentellen Bestätigung jener Theorie wäre.

5.1.5.1. Supersymmetrische Partner der Elementarteilchen. Unter Supersymmetrie versteht man eine Symmetrie zwischen Bosonen und Fermionen. Zu jedem Teilchen mit ganzzahligem Spin wird ein Partner mit halbzahligem Spin gefordert, und umgekehrt. Dementsprechend umfassen Multipletts der Supersymmetrie sowohl Bosonen als auch Fermionen. Im Gegensatz dazu enthalten die Multipletts der üblichen inneren Symmetrien (vgl. 5.1.2.3.) jeweils nur Teilchen ein- und desselben Spins. Die Beweggründe dafür, supersymmetrische Quantenfeldtheorien in Betracht zu ziehen, sind gegenwärtig rein theoretischer Natur. In supersymmetrischen Theorien kompensieren sich die Divergenzen der störungstheoretischen Feynman-Graphen (vgl. 5.1.1.8.) weitgehend. Das ist insbesondere wichtig für supersymmetrische Feldtheorien der Gravitation (Supergravitation), die sich zumindest in den niedrigsten störungstheoretischen Korrekturen als divergenzfrei erweisen, wohingegen die übliche Gravitationstheorie sogar nicht renormierbar ist. Überdies liefert die Analyse der supersymmetrischen Multipletts detaillierte Voraussagen über neue Teilchen.
Es lassen sich nämlich keinesfalls aus den bereits bekannten Teilchen Supermultipletts bilden. Vielmehr fordert die Supersymmetrie in ihrer einfachsten Variante zu jedem etablierten Teilchen ein neues, seinen Supersymmetriepartner, dessen Spin eindeutig festgelegt ist. Beispielsweise hat in der supersymmetrischen Quantenelektrodynamik jedes der Leptonen e, μ und τ einen Partner mit dem Spin 0, die sogenannten skalaren Leptonen $\tilde{e}$, $\tilde{\mu}$ und $\tilde{\tau}$. Der Partner des Photons mit dem Spin 0 wird als *Photino* bezeichnet. Dementsprechend umfaßt die supersymmetrische Quantenchromodynamik neben Quarks und Gluonen auch skalare Quarks (Spin 0) und *Gluinos* (Spin 1/2). Zu den Gravitonen (Spin 2), den theoretisch geforderten Feldquanten der Gravitation, kommt das *Gravitino* (Spin 3/2) hinzu.
Es liegt auf der Hand, daß die Supersymmetrie bestenfalls in gebrochener Form verwirklicht sein kann. Insbesondere müssen die Massen der skalaren Leptonen und des Photinos relativ groß sein, da die bis 1984 durchgeführten Experimente (vor allem die an den Elektron-Positron-Speicherringen) die übliche Quantenelektrodynamik gut bestätigen. Somit haben sich skalare Leptonen und Photinos bisher weder als reelle noch als virtuelle Teilchen (vgl. 5.1.1.8.) bemerkbar gemacht. Die Masse des skalaren Elektrons sollte demnach größer als 30 GeV/c^2 sein. Ähnliche Abschätzungen ergeben sich für das Photino, aber auch für skalare Quarks und Gluinos.
Ein besonderes Interesse verdienen supersymmetrische große einheitliche Theorien. Sie machen sehr spezifische Voraussagen hinsichtlich des Protonzerfalls, so daß es durchaus denkbar wäre, daß von dort Hinweise auf das Vorliegen der Supersymmetrie kommen, ehe noch supersymmetrische Partner entdeckt werden.

5.1.5.2. Magnetische Monopole. Fast alle großen einheitlichen Theorien lassen die Existenz von magnetischen Monopolen zu. Das sind Teilchen, die eine isolierte magnetische Ladung tragen im Gegensatz zu allen Leptonen und Baryonen, die sich wie magnetische Dipole verhalten, also die magnetische Ladung Null tragen. Die Theorie gibt für die Masse der magnetischen Monopole Werte in der Größenordnung 10^{16} GeV/c^2 an. Es besteht somit keine Aussicht, sie in absehbarer Zukunft an Beschleunigern zu erzeugen. Anders als die schweren intermediären Bosonen müssen die magnetischen Monopole stabil sein. Es könnten also durchaus Monopole als Relikte aus einer Frühphase des Kosmos übriggeblieben sein.
Nach Monopolen wurde ursprünglich mit Hilfe von Nachweisgeräten gesucht, die die Änderung des magnetischen Kraftflusses registrieren, der bei Durchgang eines Monopols durch eine Spule induziert wird. Um Änderungen des Kraftflusses auszuschließen, die durch Schwankungen äußerer Magnetfelder bedingt sind, befindet sich die Induktionsspule im Innern einer supraleitenden Metallflasche. Eine andere Nachweismöglichkeit beruht auf dem kürzlich theoretisch entdeckten Effekt des katalytischen Protonzerfalls. Protonen zerfallen im Felde eines Monopols, ohne daß dabei ein virtuelles schweres intermediäres Boson oder Higgs-Teilchen ausgetauscht werden müßten. Das Monopol bleibt in der Zerfallsreaktion $p + \text{Monopol} \rightarrow e^+ + \pi^0 + \text{Monopol}$ erhalten, deshalb spricht man von einem katalytischen Protonzerfall. Er erfolgt mit einer sehr viel größeren Wahrscheinlichkeit als der normale spontane Zerfallsprozeß. Ein Monopol, das durch eine Meßapparatur zum Nachweis des Protonzerfalls (vgl. 5.1.4.3.) hindurchfliegt, würde längs seiner Flugbahn eine Reihe von Protonzerfallsprozessen anregen, die sich von den selteneren spontanen Zerfällen unterscheiden lassen. Diese Meßapparaturen sind demnach auch zum Nachweis von Monopolen geeignet. Ein eindeutiger Beleg für die Existenz von Monopolen steht noch aus; die Experimente erlauben nur die Angabe oberer Grenzen für den mittleren Fluß der Monopole. Es ist bemerkenswert, daß die großen einheitlichen Theorien im Bund mit den Grundgleichungen der Kosmologie imstande sind, die geringe Dichte von Monopolen zu erklären. Der hierfür wesentliche Punkt ist, daß der Kosmos während eines Phasenübergangs in einen Zustand der gebrochenen Symmetrie (vgl. 5.1.3.8.) eine Periode rapider Expansionen durchgemacht hat, der die Dichte der ursprünglich gebildeten Monopole unvergleichlich stärker herabgesetzt hat, als es die normal ablaufende Expansion vermocht hätte.

5.1.5.3. Subkomponenten-Teilchen. Die Vorstellung, daß Leptonen, Quarks und vielleicht sogar die intermediären Bosonen nicht fundamental seien, sondern aus irgendwelchen Bestandteilen zusammengesetzt, ist zwar naheliegend, zugleich aber auch hypothetisch. In Erinnerung daran, daß die Periodizität in den Strukturen der Atome dadurch zu deuten war, daß die Atome als gebundene Systeme aus Kern und Elektronen behandelt wurden, bietet sich in Hinblick auf die bei den Fermionenfamilien zu beobachtende Periodizität ein analoges Vorgehen an. Es fehlt jedoch immer noch der mathematische Apparat, um solche Vorstellungen im Rahmen der Quantenfeldtheorie quantitativ, d. h. bis hin zur Berechnung der Massen und der Kopplungskonstanten von Leptonen und Quarks, zu realisieren. Die Vorstellungen von den Subkomponenten bleiben insofern aktuell, als in den zukünftigen Experimenten in jedem Fall nach entsprechenden Signalen gesucht werden wird. Das betrifft insbesondere rein leptonische Prozesse an den Elektron-Positron-Speicherringen, aber auch die tief-inelastische Lepton-Hadron-Streuung an den Beschleunigern der neunziger Jahre.
Es kann kein Zweifel daran bestehen, daß die weitere Forschung auf dem Gebiet der Elementarteilchenphysik neue Strukturen enthüllen wird, ob sie nun den bereits entwickelten Theorien entsprechen oder nicht. Das wichtigste Ziel jedoch wird sein, zu klären, ob die Gesetzmäßigkeiten auch der tieferen Schichten der Mikrowelt weiterhin mit den Prinzipien der relativistischen Quantenphysik vereinbar sind.
In diesem Zusammenhang ist auch an die Vereinbarkeit von Quantenphysik und Gravitation zu denken. Speziell im

Falle starker Gravitationsfelder gibt es prinzipielle Probleme, die noch nicht gelöst werden konnten. Möglicherweise muß irgendwann sogar die Gültigkeit der fundamentalen Vorstellungen vom Raum-Zeit-Kontinuum und seiner Dimensionalität in der Mikrowelt in Frage gestellt werden.

5.2. Kernstruktur

5.2.1. Grundlegende Eigenschaften der Kerne

5.2.1.1. Ladung und Massenzahl der Kerne. Ladung. Atome sind elektrisch neutral. Entfernt man die Elektronen der Hülle, muß die verbleibende Ladung, die des Kernes, von gleicher Größe und umgekehrtem Vorzeichen sein. Das leichteste Atom ist Wasserstoff. Entfernt man das einzige Elektron, so bleibt dem Kern, dem Proton, die Größe der Ladung des Elektrons, die Elementarladung, mit positivem Vorzeichen. Träger der Kernladung sind Protonen. Es sind immer gerade so viele, wie das neutrale Atom Elektronen enthält. Die Kernladung war schon Anfang des Jahrhunderts aus der Analyse der von Atom zu Atom verschiedenen Spektren der Röntgenstrahlen (vgl. 5.6.5.2.) ermittelt worden. Die Kernladungszahl ist gleich der Ordnungszahl im Periodensystem der Elemente. Eine andere Methode, sie zu ermitteln, beruht auf der Rutherford-Streuung (vgl. 5.5.2.2.). Bei der Streuung von α-Teilchen (Heliumkerne) an Atomen erfolgt die wesentliche Wechselwirkung zwischen der Ladung des α-Teilchens und der der Atomkerne. Unter Verwendung der experimentell ermittelten Streudaten ergibt sich die Kernladung.

Massenzahl. Die Massenzahl gibt die Anzahl der Kernbausteine (Nukleonen) an. Diese ist nicht identisch mit der Protonenzahl. Im Kern existieren noch die mit den Protonen etwa gleich schweren, aber ungeladenen Nukleonen, die Neutronen. Mit einem Massenspektrometer (vgl. 5.4.5.4.) läßt sich die Atommasse sehr genau bestimmen. (Die Masse der Elektronenhülle macht weniger als 0,1 % der Masse des Kernes aus.) Die Masse des Protons beträgt $M_p = 1{,}00759\,u$, die des Neutrons $1{,}00898\,u$. u ist die zur Erleichterung des Rechnens mit Kernmassen eingeführte Masseneinheit, und ist gegeben zu 1/12 der Masse des Kohlenstoffatoms, das aus sechs Protonen und sechs Neutronen besteht (C-12-Isotop).

Die *atomare Masseneinheit* beträgt $1\,u = 1/12\,M_{\text{Atom}}$ (C-12) $= 1{,}660277 \cdot 10^{-27}$ kg. Die Masse der Kerne ist immer in etwa ein ganzzahliges Vielfaches der atomaren Masseneinheit, da sie sich aus den Massen der Protonen und Neutronen von jeweils etwa einer Masseneinheit zusammensetzen. Zur Darstellung der Kerne verwendet man die Symbolik ${}^{A}_{Z}X$ (A = Massenzahl, X = Elementsymbol, Z = Kernladungszahl). Beispielsweise bezeichnet das Symbol ${}^{206}_{82}\text{Pb}$ das Element Blei, gegeben durch die Kernladungszahl 82 (gleich der Anzahl der Protonen) und aus 206 Nukleonen (82 Protonen und 124 Neutronen). Der genaue Massenwert dieser Atomsorte Blei (Bleiisotop 206) beträgt $205{,}97442\,u$, ist also wirklich etwa gleich der Massenzahl 206. Die Massenzahl A ist auch definiert als die in atomaren Masseneinheiten gmessene, auf ganze Zahlen gerundete Masse der Kerne.

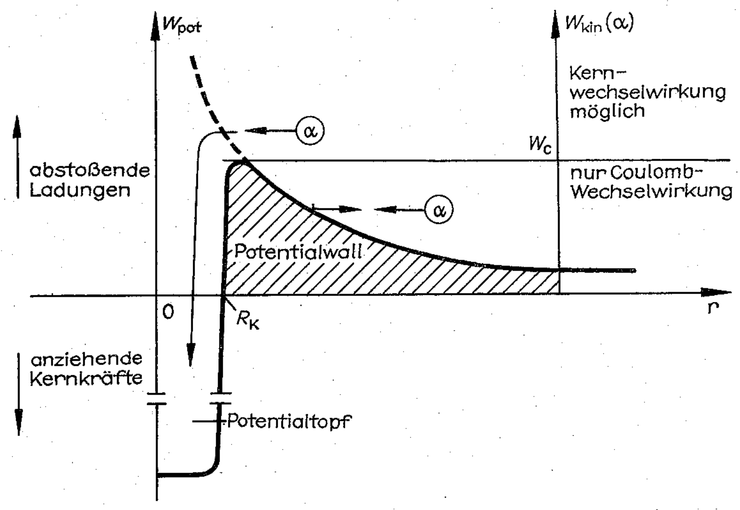


Abb. 5.2.1-1 Schematische Darstellung zur Ermittlung von Kernradien. Der Radius der Atomkerne ist durch das Einsetzen der Kernkräfte gegeben. Positiv geladene Teilchen müssen erst den durch die Coulomb-Abstoßung zustande kommenden Potentialwall überwinden ($W_{\text{kin}}(\alpha) \gtrsim W_c$), um in den Bereich der Kernkräfte zu gelangen

5.2.1.2. Radien und Form der Kerne. Es gibt viele Möglichkeiten, Kernradien genau zu bestimmen. Man kann den Kern mit α-Teilchen oder schnellen Elektronen „abtasten". Man kann seine Größe aber auch errechnen, da die sehr genau meß- und auswertbaren Anregungsspektren der Hüllelektronen den Kernradius zur Erklärung feinerer Linienverschiebungen bei verschieden großen Kernen enthalten. Am anschaulichsten erhält man Radien durch Beschuß von Kernen mit sehr schnellen Elektronen. Die Elektronen müssen Energien von $\approx 10^8$ eV aufweisen, damit ihre das Auflösungsvermögen begrenzende Wellenlänge (vgl. 5.5.4.1.) kleiner als die auszumessenden Radien wird. Wie derart Kerne vermessen werden können, soll an Hand der Abb. 5.2.1-1 erklärt werden. Eine dünne Folie mit Atomen der zu vermessenden Kernart werde mit α-Teilchen beschossen. Die Größe der Kerne ist durch das Einsetzen der

Kernkräfte definiert. Bevor die α-Teilchen in die Reichweite der Kernkräfte gelangen, unterliegen sie der Abstoßung durch den auch positiv geladenen Kern. Ab einer gewissen kinetischen Energie der α-Teilchen $W_{kin}(\alpha)$ können sie sich so weit an den Kern annähern, daß sie in den Bereich der abrupt einsetzenden, anziehenden Kernkräfte geraten. Aus der Energie, ab der das gerade möglich ist, errechnet sich der Kernradius R_K.

Quantenmechanische Betrachtungen zeigen, daß der Potentialwall im Widerspruch zur klassischen Physik nicht unbedingt überwunden werden muß. Es existiert eine gewisse Wahrscheinlichkeit des „Durchtunnelns" (vgl. 5.3.3.2.).

Wie man auch immer den Radius bestimmt, er ergibt sich in Abhängigkeit von der Massenzahl A zu etwa $R_K = 1{,}3 \cdot A^{1/3} \cdot 10^{-15}$ m. Die Einheit 10^{-15} m = 1 fm (1 Femtometer), wird meist von den Physikern – Enrico Fermi zu Ehren – als 1 Fermi bezeichnet). Mit den α-Teilchen ermittelt man die Verteilung der über die Kernkräfte wechselwirkenden Nukleonen (Protonen und Neutronen). Streut man mit Elektronen – die nicht auf Kernkräfte reagieren – erhält man die Ladungsverteilung im Kern, die Anordnung der Protonen. Wesentliche Unterschiede ergeben sich dabei nicht, die Ladungsverteilung ergibt etwas kleinere Radien, da dabei außenliegende Neutronen nicht erfaßt werden. Die Kerne sind meist von kugelförmiger Gestalt, was bisher stillschweigend angenommen wurde. Es gibt aber Massenzahlbereiche, z. B. $150 < A < 190$, wo auch stärkere Abweichungen von der Kugelform zu finden sind.

5.2.1.3. Massendichte der Kerne. Da die Elektronenmasse mit $M_e = 5{,}5 \cdot 10^{-4}\,u$ nur etwa den 1/2000sten Teil der Nukleonenmasse ausmacht, ist fast die gesamte Masse des Atoms im Kern angeordnet. Er bildet die „Sonne" im Planetensystem des Atoms und wird in großer Entfernung von den Elektronen „umkreist".

Anschaulich stellt man sich den Kern als aus dicht an dicht haftenden kugelförmigen Protonen und Neutronen gebildet vor. Diese Anschauung einer Kugel aus lauter Nukleonenkugeln bestätigt sich bei einer Betrachtung der Dichteverteilung im Kern. Die Dichte ϱ_K ergibt sich aus der Masse M_K und dem Volumen V_K zu $\varrho = M_K/V_K$. Da $M_K \approx A \cdot u$ und aus dem Kernradius $R = 1{,}3 \cdot A^{1/3}$ das Volumen $V_K = (4\pi/3)\,R_K^3$ gegeben ist, wird $\varrho_K \approx 3 \cdot A \cdot u/(4\pi \cdot 1{,}3^3 \cdot A)$, also ϱ_K = konstant$_A$. Die Formel für die *Kerndichte* enthält nicht mehr die Massenzahl. Die Kerndichte ist also (unabhängig von A) für alle Kerne konstant. Das entspricht der Vorstellung dicht gepackter und sich auch bei großer Anzahl nicht komprimierender Nukleonenkugeln. Am Rande des Kernes fällt die Dichte in einem Bereich von $\approx 2{,}4 \cdot 10^{-15}$ m von 90 % auf 10 % der Dichte ϱ_K ab. Dieser Bereich wird Kernoberfläche genannt und zeigt an, daß sich die konstante Kerndichte noch nicht bei den leichten Kernen einstellt, diese bestehen im wesentlichen aus Oberfläche.

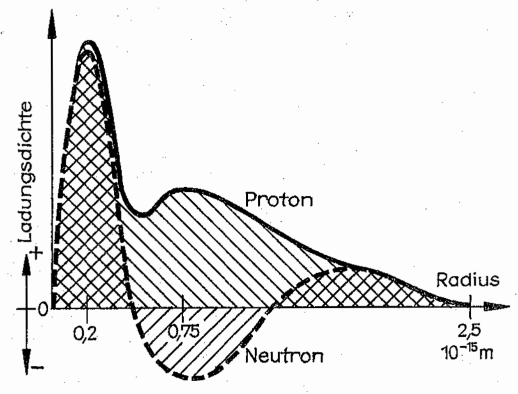


Abb. 5.2.2–2 Skizze der Ladungsdichteverteilung der Nukleonen

Der Zahlenwert für die aus der oben angegebenen Formel berechneten Kerndichte ergibt den verblüffend großen Wert von $\varrho_K \approx 2 \cdot 10^{17}$ kg m^{-3}. Um diese Dichte zu veranschaulichen, müßte man sich einen beladenen 200 000-t-Tanker auf die Größe von 1 mm^3 zusammengedrückt vorstellen. Derartige Massendichten erreichen Neutronensterne, wo durch extrem große Masseansammlungen die Gravitationskraft alle anderen Kräfte übertrifft und die Atome bis auf ihre Kernabstände aneinander gepreßt werden.

5.2.2. Aufbau der Atomkerne

5.2.2.1. Nukleonen und ihre Eigenschaften. *Nukleonen* nennt man die etwa massengleichen Bausteine der Kerne, Protonen und Neutronen. Ihre einfachsten Eigenschaften sind in Tab. 5.2.2–1 eingetragen. Die Massen der Nukleonen sind sehr genau meßbar und ergeben sich ohne jede Individualität als immer genau gleich für jedes Proton bzw. Neutron. Die Masse ist eine Elementarteilchen charakterisierende Eigenschaft. Das gleiche gilt für die Ladung. Das Proton ist Träger einer Elementarladung, gleich groß, aber entgegengesetzten Vorzeichens der des Elektrons. Das Neutron ist in elektrischen Feldern nicht aus seiner ursprünglichen Richtung ablenkbar, es zeigt sich nach außen hin ungeladen.

Tab. 5.2.2–1 Einfache Eigenschaften der Nukleonen

Nukleon	Masse	Ladung	Drehimpuls	magnetisches Moment
Neutron n	1,008 98 u	0	$1/2\,\hbar$	$-1{,}91\,\mu_K$
Proton p	1,007 59 u	+ e	$1/2\,\hbar$	$2{,}79\,\mu_K$

Beide Nukleonen haben einen Eigendrehimpuls (Spin). In der anschaulichen Teilchenvorstellung entspricht das einer Eigenrotation. $\hbar = h/2\pi$ ist die atomare Drehimpulseinheit (h = Plancksches Wirkungsquantum = $6{,}626 \cdot 10^{-34}$ Js).

Das Magnetfeld des Protons ist anschaulich verständlich. Durch die Eigendrehung des Nukleons rotiert die Ladung, was in der Wirkung der Erzeugung eines Magnetfeldes einer stromdurchflossenen Spule entspricht. Das magnetische Moment des Neutrons ist so nicht zu verstehen, da keine Ladung vorhanden ist. Das magnetische Moment wäre durch eine mit dem Spin rotierende negative Ladung erklärbar.

Ladungsdichteverteilung. 1960 verwendete Hofstadter Elektronen mit einer kinetischen Energie von 600 MeV zur genauen Größenbestimmung der Kerne. Die Energie ist ausreichend, die Ladungsverteilung in Kernen genau austasten zu können. Hofstadter schoß die Elektronen auf flüssigen Wasserstoff. Da die Atome des Wasserstoffs aus jeweils nur einem Proton bestehen, ließ sich so die Ladungsverteilung des Protons bestimmen. Danach bestimmte er die Ladungsverteilung des aus einem Proton

und einem Neutron bestehenden Deuterons ^2_1H. Aus beiden Ladungsverteilungen zusammen kann man die des Neutrons ermitteln. Abb. 5.2.2–2 zeigt das Ergebnis der Messungen. Danach haben die Nukleonen eine innere Struktur, die sich durch eine inhomogene Ladungsverteilung äußert. Auch das nach außen hin neutrale Neutron hat eine innere Ladungsverteilung, die sein magnetisches Moment erklärt.

Lebensdauer. Die Lebensdauer des Protons wurde bis vor kurzem als unendlich groß angegeben. Das gilt auch heute noch mit genügender Genauigkeit, obwohl versucht wird, eine vorhergesagte Halbwertszeit von $\approx 10^{31}$ Jahre nachzuweisen. Das im Kern gebundene Neutron ist stabil. Das freie Neutron ist instabil. Es zerfällt nach knapp 1000 Sekunden in Proton, Elektron und Antineutrino, $N \rightarrow p^+ + e^- + \bar{\nu}$. Es wäre aber falsch zu formulieren, daß das Neutron aus Proton, Elektron und Neutrino besteht; die Teilchen bilden sich erst beim Zerfall.

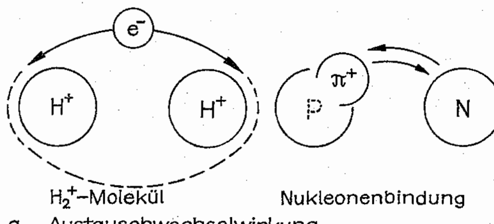
5.2.2.2. Nukleonenbindung im Kern. Die bindenden Kräfte im Kern müssen sehr groß sein, größer als die abstoßenden der Protonen untereinander (um zwei Protonen miteinander in Berührung zu bringen, benötigt man eine Kraft von ≈ 27 N). Diese Überlegung führt schon zu Bindungsenergien in der Größenordnung von Megaelektronenvolt. Tatsächlich sind die Bindungsenergien ≈ 8 MeV je Nukleon.

Im Vergleich dazu sind die chemischen Bindungen millionenmal schwächer. *Kernkräfte* sind auch von anderer Art als Gravitations- oder Coulombsche Kräfte. Letztere reichen unendlich weit, wenn sie auch mit dem Quadrat des Abstandes (mit $1/r^2$) schwächer werden. Kernkräfte sind sehr kurzreichweitig. Noch kurz bevor sich Nukleonen berühren, ist keine Wirkung zu bemerken, bei etwas weiterer Annäherung (Berührung) sind die Kräfte voll ausgeprägt, etwa wie eine Klebe(ver)bindung. Der Verlauf der Kernkräfte in Abhängigkeit vom Abstand r der Nukleonen ist nicht genau bekannt. Das Potentialtopfmodell – ein Potential nach Abb. 5.2.1–1 verwendend – beschreibt den Sachverhalt gut. Eine einfache Näherung ist das *Kastenpotential* $V_K =$ konst. für $r \leqq r_0$, $V_K = 0$ für $r > r_0$.

Kernkräfte wirken gleichstark zwischen den Protonen und zwischen den Neutronen, wie auch zwischen Protonen und Neutronen. Die Kräfte wirken stark anziehend (bindend), haben aber auch eine abstoßende Komponente, weil Nukleonen sich nicht komprimieren, nicht durchdringen. Die Nukleonen bleiben immer auf gleichem Abstand untereinander, erklärbar durch abstoßende Kräfte.

Eine weitere Eigenschaft der Kernkräfte ist es, sich abzusättigen. Nukleonen sind paarig im Kern recht stabil. Alle Kerne haben etwa gleich viele Protonen wie Neutronen. Außerdem sind in der Elementhäufigkeit und bei anderen Eigenschaften, die die Stabilität charakterisieren, Elemente mit geraden Protonen- oder/und Neutronenzahlen bevorzugt. Das weist auf eine paarweise Absättigung der Bindungskräfte. Für den Sättigungscharakter der Kernkräfte gibt es eine Analogie bei den chemischen Bindungen – die homöopolare Bindung und ihre Erklärung über Austauschwechselwirkungen. Das führt zugleich zu einer Beschreibungsmöglichkeit für Kernkräfte.

Abb. 5.2.2–3 versinnbildlicht Erklärung und Nachweis der Austauschwechselwirkung. In der Physik der Atome erklärt man die Existenz des aus einem Wasserstoffatom und einem Wasserstoffion bestehenden H_2^+-Moleküls mit dem ständigen Wechsel des einzigen Elektrons von einem Atom zum anderen. Atom und Ion wechseln so ständig ihre Identität. Diese Austauschwechselwirkung ergibt eine bindende Kraft.



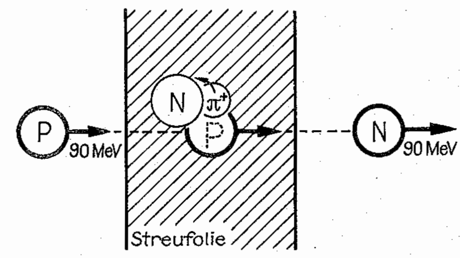


Abb. 5.2.2–3 (a) Erklärung der Nukleonenbindung durch π-Mesonenaustausch analog der H_2^+-Molekülbindung durch Elektronenaustausch, (b) Nachweis der Existenz des π-Mesons durch seinen Wechsel von Strahlteilchen zu einem Foliennukleon

Die klammernden Teilchen zwischen den Nukleonen wurden von YUKAWA 1935 vorhergesagt. Ein Jahrzehnt später konnten sie experimentell nachgewiesen werden. Es ist ein Mesonentriplett, bestehend aus den π^0-, π^-- und π^+-Mesonen.

Die Masse beträgt, wie auch vorhergesagt, ≈ 270 Elektronenmassen, der Spin ist 0. Die Lebensdauer des freien positiv geladenen π^+-Mesons und seines Antiteilchens, des negativ geladenen π-Mesons, beträgt $2{,}6 \cdot 10^{-8}$ s, die des neutralen π^0-Mesons $8{,}3 \cdot 10^{-17}$ s. Das die Austauschwechselwirkung vermittelnde π-Meson ist im Kern nicht als freies Teilchen nachweisbar, man sagt, es existiert nur virtuell. Dennoch gibt es einen anschaulichen Beweis seiner Existenz durch ein Streuexperiment. Läßt man einen Protonenstrahl hoher Energie eine dünne Folie durchdringen, so erscheint mitunter hinter der Folie ein Neutron genau in Richtung des Protonenstrahles mit genau der Energie der Protonen. Nach Energie- und Impulssatz kann es nicht aus einem Atom der Folie herausgeschlagen worden sein. Die Erklärung ist folgende: Ein Strahlproton kommt so dicht an einem Neutron vorbei, daß, wie im Kern, ein π^+-Meson vom Proton zum Neutron wechselt, wodurch das Protor als Neutron weiterfliegt und das Neutron in der Folie zum Proton wird (vgl. Abb. 5.2.2–3). Das Experiment geht genauso mit einem Neutronenstrahl. Ein π^+-Meson wechselt aus einem Proton der Streufolie in ein Strahlneutron, das dann als Proton die Folie verläßt.

5.2.2.3. Systematik der Kerne. Das Periodensystem der Elemente (vgl. 5.6.3.) bezieht sich auf die chemischen Eigenschaften der Atome, die aus dem Aufbau der Elektronenhülle resultieren. Eine Vielfalt mit steigendem Atomgewicht periodisch wiederkehrender Eigenschaften sind zu berücksichtigen. Eine Systematisierung der Elemente ist in

Zeilen, Spalten, Untergruppen und weiteren Feinheiten einer zweidimensionalen Darstellung ergiebig.

Eine Systematisierung der Kerne ist wesentlich einfacher. Man macht dies in Form einer Nuklidkarte. Das ist die Darstellung aller möglichen Kerne in einer Ebene. Auf die Ordinate und Abzisse wird die Zahl der Protonen bzw. der Neutronen aufgetragen (vgl. Abb. 5.3.3–1).

Für jeden Kern ergibt sich ein Rechteckfeld, das alle wesentlichen Daten des Nuklides enthält: Lebensdauer, bzw. wenn stabil, dann Anteil dieses Isotopes an der Menge aller stabiler Isotope des Elementes, Atomgewicht, Wirkungsquerschnitt für spezielle Reaktionen, Zerfallsart und Energie der Strahlung. Nuklidkarten sind in der Kernphysik gebräuchlich.

Elemente mit gleicher Protonenzahl und verschiedener Neutronenzahl werden *Isotope* genannt. Ihre Bedeutung ergibt sich aus dem chemisch gleichen Verhalten aller Isotope eines Elementes. *Isotone* sind Kerne mit gleicher Neutronenzahl und verschiedener Protonenzahl. Gleichschwere Kerne verschiedener Protonen- und Neutronenzahl werden als *Isobare* bezeichnet.

Isotopieeffekte. Isotope eines Elementes haben aufgrund der unterschiedlichen Neutronenzahl verschiedene Massen. Zudem sind sie noch von etwas verschiedener Größe. Die größten Massenunterschiede treten bei leichten Elementen auf. Deuterium ist etwa doppelt so schwer wie Wasserstoff. Isotope können in ihrem kernphysikalischen Verhalten sehr verschieden sein. Wasserstoff (H-1) und das zu 0,015% in der natürlichen Zusammensetzung enthaltene Deuterium (H-2) sind stabil. Das zur radioaktiven Markierung von organischen Substanzen und zur Kernfusion verwendete Tritium (H-3) hat nur eine Halbwertszeit von 12,3 Jahren. Deuterium hat einen sehr geringen Neutronen-Einfangquerschnitt und dient in der Form von schwerem Wasser (D_2O) als Moderatorsubstanz in Kernreaktoren. Die verschiedenen Massen der Isotope führen zu verschieden starker Ablenkung in elektrischen und magnetischen Feldern. Isotope Elemente zeigen im gasförmigen Zustand verschiedene mittlere Geschwindigkeiten und damit verschiedene Diffusionsverhalten. Das chemische Verhalten der Isotope ist kaum unterscheidbar. Beispielsweise wird der Radiokohlenstoff (C-14) bei der Bildung organischer Substanzen genauso eingebaut wie das stabile Isotop C-12.

Jedoch sind auch in der Natur durch eine große Zyklenzahl von Reaktionsabläufen bzw. über lange Zeiträume Isotopieeffekte beobachtbar. Durch den Wasserkreislauf der Natur sind die schwereren Isotope H-2 bzw. O-18 der Atome des Wassers in den Ozeanen etwas angereichert bzw. im Grundwasser in geringerer Menge vorhanden.

5.2.2.4. Isotopentrennung. Isotope eines Elementes können durch chemische Methoden nicht getrennt werden. Die Trennung ist aber wesentlich, z. B. um den Reaktorbrennstoff U-235, der nur in sehr geringer Menge (0,7%) im U-238 enthalten ist, in reiner Form oder auch nur angereichert zu bekommen. Aus dem Massenunterschied von 1,3% zwischen U-235 und U-238 errechnet sich ein Unterschied von 0,4% in der mittleren Geschwindigkeit der entsprechenden Moleküle im gasförmigen Zustand. Dadurch ergibt sich bei Diffusion durch eine poröse Wand ein Vorteil für die leichtere Komponente, die etwas leichter durchkommt. Wiederholt man diesen Vorgang sehr oft, reichert sich das U-235 mehr und mehr an.

Isotopentrennanlagen dieser Art haben gewaltige Ausmaße und verbrauchen eine Energiemenge, wie sie zur Versorgung einer Großstadt benötigt wird. Ein modernes Gerät zur Isotopentrennung ist die Ultrazentrifuge. In extrem schnell drehenden Trommeln kann eine Zentrifugalbeschleunigung erzeugt werden, die die Erdbeschleunigung millionenmal übertrifft. Dadurch werden die durch die verschiedenen Isotopenmassen wirkenden Kraftunterschiede schon merklich, womit eine gute Trennmöglichkeit gegeben ist. Mitunter sind andere Eigenschaften zur Isotopentrennung nutzbar. So ist der Siedepunkt des normalen Wassers 100,0 °C, der des schweren Wassers 101,4 °C (statt $^1_1$H enthält es $^2_1$D). Durch vorsichtiges Sieden kann man diese isotopen Moleküle voneinander trennen.

Genaueste Methode zur Isotopentrennung ist die Massenspektrometrie (vgl. 5.4.5.4.). Das ist eine sehr empfindliche Methode zur Massebestimmung, die aber nicht produktiv für größere Mengen eingesetzt werden kann. Neueste Variante zur Isotopentrennung ist die Anregung bestimmter Energieniveaus der Elektronenhülle, die von Isotop zu Isotop gering differieren, mit genau abgestimmten leistungsstarken Laseranlagen. Das sensibilisiert die Isotope z. B. für bestimmte chemische Reaktionen. Mitunter kann man auch genau nur ein Isotop ionisieren, was dann als Ion leicht mittels eines elektrischen Feldes abzutrennen ist.

5.2.3. Bindungsenergiebetrachtungen

5.2.3.1. Masse-Energie-Äquivalenz. Einstein formulierte eine Beziehung allgemeiner Gültigkeit zwischen Masse M und Energie W, $W = Mc^2$ (c = Lichtgeschwindigkeit). Diese Beziehung ist auch als Reaktionsgleichung zu schreiben, $W \leftrightarrow Mc^2$.

Sie besagt, daß Energie in Masse wandelbar ist und Masse in Energie. Ein Teilchen kann auf Kosten seiner Masse Energie in Form von Licht oder γ-Quanten abgeben. Zwei Teilchen, z. B. Elektron und Positron (generell immer Teilchen und Antiteilchen), können zerstrahlen, d. h. als Masseteilchen verschwinden und in die Energie zweier γ-Quanten übergehen. Ein γ-Quant ausreichender Energie kann unter Verlust seiner Existenz ein Teilchen-Antiteilchen-Paar erzeugen.

Damit ist aus der Sicht der klassischen Physik der Satz von der Erhaltung der Energie (Energiesatz) und auch der der Erhaltung der Masse verletzt. Man kann beide Sätze formal retten, wenn jeweils die Massen- bzw. Energiedifferenz mittels Einsteins Masse-Energie-Äquivalenz aus der Energie- bzw. Massenbilanz hinzu gerechnet wird. Es ist aber günstiger, Energie- und Massenerhaltungssatz in ihrer voneinander unabhängigen Gültigkeit aufzugeben, und dafür den *Erhaltungssatz der Materie* einzuführen:

1. Materie kann weder gebildet, noch vernichtet werden.
2. Daseinsformen der Materie sind Energie und Masse; sie können sich ineinander umwandeln.

5.2.3.2. Bindungsenergie und Massendefekt. Bindungsenergie. Atomkerne sind sehr stabile Gebilde; die Nukleonen sind fest miteinander verbunden, im Kern gebunden. Die Bindungsenergie eines Nukleons ist die Energie, die mindestens aufzubringen ist, um das Nukleon vom Kern zu trennen.

Auf der Energieskala ist Bindungsenergie negative Energie. Sie ist erst aufzubieten, zu kompensieren, ehe kinetische Energie zustande kommen kann. Die Bindungsenergie stabiler Kerne liegt im Mittel bei 8 MeV je Nukleon. 8 MeV Energie entsprechen ≈15,6 Elektronenmassen oder 0,85% der Masse eines Nukleons. Bei schweren Kernen ist somit die gesamte Bindungsenergie der Masse von ein bis zwei

Nukleonen äquivalent. Beispielsweise hat $^{197}_{79}Au$ eine Masse von 196,965 u; die Summe der Massen der 79 Protonen ($79 \cdot 1{,}00759\,u$) und der 118 Neutronen ($118 \cdot 1{,}00898\,u$) ergibt jedoch 198,659 u. Somit ist die Differenz zwischen Masse des Kernes und Masse der einzelnen Bausteine fast 1,7 Masseeinheiten. Die Masse des Kernes ist geringer als die seiner Bausteine. Derart ist Bindungsenergie zu verstehen: Die Zerlegung in einzelne Bausteine ist ohne Zuführung von Energie – um die Massedifferenz aufzubringen – nicht möglich.

Massendefekt. Die Massedifferenz Δm wird *Massendefekt* genannt, $\Delta m = Z \cdot m_p + (A - Z)\,m_N - m_{Kern}$. Er ist die Differenz aus der Masse der Z Protonen und der $(A - Z)$-Neutronen zur Masse des Kernes. Der Massendefekt entspricht der Bindungsenergie des Kernes, $\Delta mc^2 = -W_B$.

Der Massendefekt existiert in gleicher Weise bei chemischen Reaktionen. Nur ist z. B. bei der Verbrennung von Kohle der Massendefekt $\approx 10^{-6}$mal geringer als bei Kernreaktionen. Genaue Messungen weisen aber genau den sich entsprechend der etwas verstärkten Bindung ergebenden Massendefekt aus.

5.2.3.3. Möglichkeiten zur Kernenergienutzung. Abb. 5.2.3–1 zeigt den Verlauf der Bindungsenergie je Nukleon in Abhängigkeit von der Nukleonenzahl. Er weist aus, daß bei mittelschweren Kernen die stärkste Bindung erreicht wird, und daß sowohl leichte als auch schwere Kerne weniger stark gebunden sind. Die Abbildung erklärt die Freisetzung von Energie bei der Verschmelzung leichter Kerne (Fusion) und der Spaltung schwerer Kerne. Dies sei am Beispiel der Kernspaltung erläutert. Bei der *Kernspaltung* sind die Nukleonen in den beiden Teilen nachher fester gebunden als zuvor im schweren Kern. Die Bindung verstärkt sich, die Bindungsenergiedifferenz wird an die Umwelt abgegeben. Der Unterschied in der Bindung je Nukleon beträgt ≈ 1 MeV. Kerne mit Massenzahlen um $A = 200$ liefern nach der Spaltung in zwei Teilstücke mit Massenzahlen um $A = 100$ eine Energie von 200 mal 1 MeV, da danach jedes der Nukleonen um 1 MeV fester gebunden ist. Diese freiwerdende Energie ist millionenmal größer als bei Prozessen in der Elektronenhülle. Bei der Verbrennung von Kohle wird je Atom eine Bindungsenergiedifferenz in der Größenordnung von Elektronenvolt frei.

Kernfusion ist die Verschmelzung leichter Kerne zu schwereren. Der Prozeß ist aber prinzipiell schwieriger zu realisieren als die Spaltung, da erst Abstoßungskräfte der beiden positiv geladenen leichten Kerne überwunden werden müssen, um in den Bereich der Kernkräfte zu gelangen.

Die bei der Fusion freiwerdende Energie ist enorm. Könnte man einen Heliumkern aus zwei Protonen und zwei Neutronen zusammensetzen, würden 28 MeV Energie an die Umwelt abgegeben. Das entspricht der Umsetzung von 3% der Masse eines Nukleons in Energie.

5.2.3.4. Das Tröpfchenmodell zur Erklärung des Verlaufes der Bindungsenergie. Kernmodelle. Man bedient sich einer Modellvorstellung, um daran Vorgänge besser verstehen zu können, Rechenansätze zu finden oder einen Zusammenhang in der Analogiebetrachtung zu veranschaulichen. Zur Erklärung des Verhaltens der Kerne kommt man leider nicht mit einem Modell aus (Kernkräfte sind nicht exakt bekannt, und da ein Kern in der Regel mehrere Nukleonen enthält, ist immer das mathematisch schwierige Mehrkörperproblem zu behandeln).

Viele Modelle werden verwendet, ein „optisches“ Modell für spezielle Kernreaktionen, ein „Schalen“-Modell für bestimmte Kerneigenschaften oder das nun zu diskutierende *Tröpfchenmodell.*

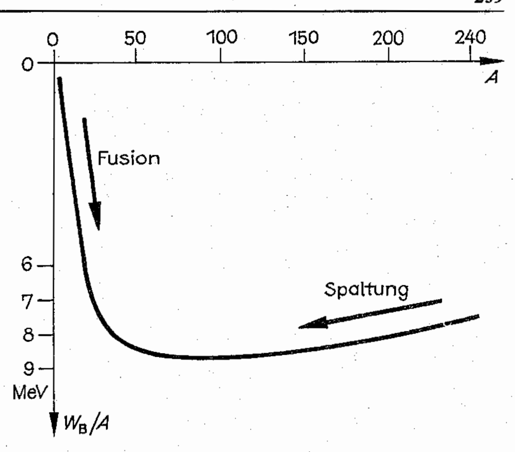


Abb. 5.2.3–1 Darstellung der Bindungsenergie je Nukleon als Funktion der Massenzahl und – durch Pfeile gekennzeichnet – der Möglichkeiten, Kernreaktionen mit positiver Energiebilanz vorzufinden

Bethe-Weizsäcker-Formel. Der Kern weist Ähnlichkeiten mit einem Flüssigkeitstropfen auf. In einem Flüssigkeitstropfen ist wie im Kern die Dichte unabhängig von seiner Größe konstant. Die zwischenmolekularen Kräfte sind wie die Kernkräfte kurzreichweitig und haben Sättigungscharakter. Ein Flüssigkeitstropfen hat auch wie der Kern eine kugelförmige Oberfläche. Die Physiker H. BETHE und C. F. WEIZSÄCKER stellten 1935 eine Formel zur Beschreibung der Bindungsenergie in ihrer Abhängigkeit von der Massenzahl A auf. Sie gingen zum Verständnis der Kräfte von den Vorstellungen in einem Flüssigkeitstropfen aus. Ihre Formel gibt sehr genau den Verlauf der Bindungsenergie, wie er in Abb. 5.2.3–1 dargestellt ist, wieder. Nach BETHE und WEIZSÄCKER ist die totale Kernbindungsenergie W_B durch fünf additive Terme erklärbar: $W_B = -aA + b \cdot A^{2/3} + c(Z^2/A^{1/3}) + d(N - Z)^2/A + e$ (aA = Volumenterm, $b \cdot A^{2/3}$ = Oberflächenterm, $c \cdot Z^2/A^{1/3}$ = Abstoßungsterm, $d(N - Z)^2/A$ = Symmetrieterm, e = Paarungsterm).

Die Bindungsenergie ist, nach allem bisher Gesagten, proportional der Zahl der Nukleonen A. Das wird durch den ersten, den Volumenterm berücksichtigt. Alle anderen Terme verringern die Bindungsenergie aus verschiedenen Gründen. Der zweite Term kennzeichnet die Wirkung der Oberfläche. Die Kerne an der Oberfläche sind lockerer gebunden und verringern die Bindung. Der Term ist proportional der Größe der Oberfläche ($\sim R^2$, d. h., $\sim A^{1/3 \cdot 2} = A^{2/3}$). Kern und Wassertropfen vergrößern ihre Stabilität (Bindung) durch Verringerung der Oberfläche, d. h., sie nehmen eine Kugelgestalt an. Der dritte Term hat keine Analogie zum Flüssigkeitstropfen. Der Term berücksichtigt die Abstoßung der Protonen untereinander und das lockert ebenfalls die Bindung. Daß keine stabilen Kerne mit $Z > 82$ existieren, liegt an der dort schon zu starken Protonenabstoßung. Der vierte, der Symmetrieterm, berücksichtigt die besondere Stabilität für $N = Z$, ein Effekt, der erst quantentheoretisch begründet werden kann. Das gleiche gilt für den Paarungsterm, mittels dem die Bindungsenergie um einen konstanten Betrag für gg-Kerne (gerade Protonen- und Neutronen-

zahl) erhöht und um den gleichen Betrag für *uu*-Kerne verringert wird (ungerade Protonen- und Neutronenzahl). Der Term ist Null für alle anderen Kerne.

Die Formel ist modellgestützt entstanden, – man sagt halbempirisch – ihre genaue Anpassung an die gemessenen Bindungsenergiewerte erfolgt mit den vor den einzelnen Termen stehenden Faktoren *a, b, c, d* und *e*. Eine gleich gute Anpassung der Formel an die realen Bindungsenergien wird durch mehrere Sätze von Werten für diese Faktoren erreicht. Hinreichend gut ist z. B. folgender Satz:

$a = 14{,}1\,\text{MeV}$; $b = 13{,}1\,\text{MeV}$; $c = 0{,}595\,\text{MeV}$; $d = 18{,}1\,\text{MeV}$ und $e = 33{,}5\,\text{MeV}/A^{3/4}$.

Mit dieser Formel für die Bindungsenergien sind die experimentellen Werte ab $A \geqq 15$ mit einem Fehler von 1% errechenbar. Einiges kann die Formel nicht beschreiben. Für kleine Nukleonenzahlen ist klar, daß nicht genügend Nukleonen vorhanden sind, um eine annähernd glatte Oberfläche zu bilden, wie es die Vorstellung vom Wassertröpfchen erfordert. Hier liefert der Oberflächenterm falsche Werte.

5.2.3.5. Bindungsenergien isobarer Kerne. Eine Beschreibung der Bindung nur in Abhängigkeit von der Massenzahl *A* ist unvollständig.

Für jede Massenzahl ist derart nur ein Nuklid betrachtet. Eine reale Darstellung der Bindungsenergie muß zu jedem *A* mehrere Protonen- und Nukleonenzahlen zulassen.

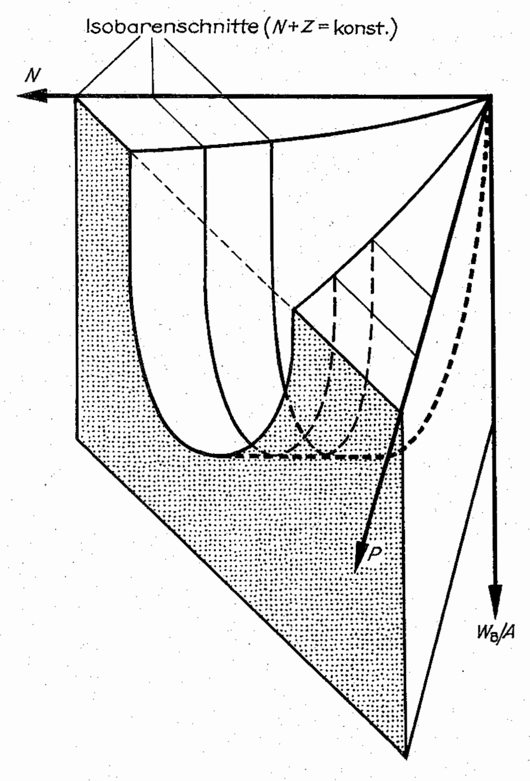


Abb. 5.2.3–2 Das Bindungsenergietal und ein Schnitt in die Ebene gleichschwerer (isobarer) Kerne. Die Bindungsenergiekurve nach Abb. 5.2.3–1 verläuft entlang der Talsohle (rechte gestrichelte Linie) (N = Zahl der Neutronen, P = Zahl der Protonen)

Abb. 5.2.3–1 läßt sich in eine dreidimensionale Darstellung vertiefen, in dem die *A*-Achse zur *N-P*-Fläche erweitert wird (Abb. 5.2.3–2).

Wichtig sind die Bindungsenergien für konstante *A*, die *Isobarenschnitte*. Der stabilste Punkt – das Bindungsenergietal – kann durch β^--Zerfälle ($N \rightarrow P$) und β^+-Zerfälle ($P \rightarrow N$) erreicht werden. Dort liegen die stabilen Kerne. Genauere Betrachtungen liefern folgende Regeln für isobare Kerne:

1. Kerne mit ungerader Massenzahl haben höchstens einen stabilen Kern.
2. Bei Kernen gerader Massenzahl existieren meistens zwei stabile Kerne.

5.2.4. Weitere Eigenschaften der Kerne

5.2.4.1. Kernspin, magnetische und elektrische Momente.

Kernspin. Nukleonen haben den Eigendrehimpuls (Spin) $I = 1/2\hbar$. Die Einheit der Drehimpulsquantelung $\hbar = h/2\pi$ wird in verkürzter Schreib- und Sprechweise nicht erwähnt, man gibt nur noch die Vielfachen der Einheit an, die Spinquantenzahl s, für Protonen und Neutronen ist also $s = 1/2$. Auch die Rotation eines Nukleons, etwa um den Schwerpunkt des Kernes, ein Bahndrehimpuls, ist existent. Für den Bahndrehimpuls im Kern versagt die mechanistische Anschauung eines ungehinderten Bahnumlaufes. Nach dem Tröpfchenmodell sind die Nukleonen dicht an dicht liegend und stark an den nächsten Nachbarn gebunden. Die Bahndrehimpulsquantenzahl l eines Nukleons kann die Werte $l = 0, 1, 2, 3 \ldots$ annehmen.

Bahn- und Eigendrehimpuls eines Nukleons addieren sich zum Gesamtdrehimpuls $j = l + 1/2$ oder $l - 1/2$, j ist somit immer halbzahlig. Der Gesamtdrehimpuls j eines Kernes bildet sich aus der Summe aller Nukleonen-Gesamtdrehimpulse, $J = \sum j_i$.

Somit ist bei gerader Nukleonenzahl J stets ganzzahlig und bei ungerader Nukleonenzahl stets halbzahlig. Nukleonen binden sich miteinander zumeist unter Kompensation des Drehimpulses, so daß keine größeren Kerndrehimpulswerte auftreten. Insbesondere sind die Drehimpulswerte von gg-Kernen stets Null. Alles bisher zum Kerndrehimpuls Gesagte betrifft den Grundzustand. Kerne können angeregte Zustände bilden, in denen dann auch höhere Drehimpulse zustande kommen.

Magnetische Kernmomente. Die Rotation von Ladungen erzeugt einen magnetischen Dipol, ein magnetisches Moment. Kernmagnetische Momente haben nur $\approx 1/2000$ der Größe der magnetischen Momente der Atomhülle.

Das magnetische Moment M ergibt sich aus der elektrischen Ladung q und dem Drehimpuls S zu $M = S \cdot q/(m \cdot c)$.

Die Masse m im Nenner bewirkt die Kleinheit des kernmagnetischen Momentes gegenüber dem des Elektrons. Die Protonenmasse ist 1836 mal größer als die des Elektrons, wodurch das magnetische Moment um diesen Faktor beim Proton kleiner ist. In Realität sind die Zusammenhänge komplizierter, aber größenordnungsmäßig stimmt es so.

Kernspin und magnetische Momente sind als Hyperfeinstruktur der Spektrallinien experimentell zugänglich. Der Kerndrehimpuls addiert sich vektoriell zum Drehimpuls der Hülle. Die verschiedenen Einstellmöglichkeiten des Kerndrehimpulses zu dem der Hülle bewirken die Hyperfeinstruktur der Spektrallinien (vgl. 5.6.2.3.).

Bei zweiatomigen Molekülen kopppeln die Kernspins der Atome in einfacher Weise miteinander. Beim H_2-Molekül können die Spins der beiden Protonen zueinander parallel stehen und sich addieren oder antiparallel stehen und sich aufheben.

Elektrische Kernmomente. Der Kern kann kein elektrisches Dipolelement besitzen, da er nur Ladungen eines Vorzeichens (Protonenladungen) enthält. Weicht die Form des Kernes von der einer Kugel ab, ist sie etwa ein Rotationsellipsoid, so resultiert daraus ein Quadrupolmoment. Eine genaue Analyse der Hyperfeinstrukturaufspaltung führt zur Bestimmung von Art und Größe elektrischer Kernmomente.

5.2.4.2. Das Schalenmodell der Atomkerne. Verschiedene Eigenschaften der Atome, die nicht durch Effekte der Elektronenhülle erklärbar sind, lassen es als notwendig erscheinen, bestimmte Nukleonenzahlen im Kern auszuzeichnen. So muß es eine Eigenschaft der Kerne sein, daß die Elemente ^4_2He, $^{16}_8O$ und $^{40}_{20}Ca$ besonders häufig im Weltall (von allen existierenden Elementen überhaupt und nicht nur auf der Erde) auftreten. Desweiteren existieren besonders viele stabile Atome mit den Neutronenzahlen 20, 28, 50 und 82. Auch findet man bei den Protonenzahlen 50 und 82 besonders viele Isotope.

Magische Zahlen. Derart und in anderer Weise ausgezeichnete Nukleonenzahlen beziehen sich gleichermaßen auf Protonen wie auf Neutronen und sind im Sprachgebrauch als magische Zahlen eingegangen. Es sind die Zahlen 2, 8, 20, 28, 50, 82, 126. Damit sind ^4_2He, $^{16}_8O$ und $^{40}_{20}Ca$ Elemente mit magischen Protonen- und magischen Neutronenzahlen, genauso das häufigste Bleiisotop $^{208}_{82}Pb$.

Eine Theorie, die versucht, diese Werte zu errechnen, führt noch zu der Protonenzahl 114 und der Neutronenzahl 184. Derart schwere Kerne sind nicht stabil, aber man vermutet, daß unter all den kurzlebigen Transuranen die mit als magisch vermuteten Zahlen noch am stabilsten sind und sogenannte *Stabilitätsinseln* bilden.

Schalenmodell. Naheliegend ist der Versuch, die magischen Zahlen mit Werten für abgeschlossene Nukleonenschalen im Kern analog den abgeschlossenen Elektronenschalen im Periodensystem der Elemente zu erklären. Die Vorstellung eines zwiebelförmigen Schalenaufbaues der Kerne liefert noch die besondere Schwierigkeit, sich die Schalen für Protonen und Neutronen einzeln und sich gegeneinander nicht beeinflussend vorstellen zu müssen. Entscheidend ist aber nicht das Vorstellungsvermögen, eher schon eine vieles erklärende Beschreibungsmöglichkeit. Der Aufbau der Elektronenschalen ergibt sich aus den vier Quantenzahlen und dem Pauli-Prinzip dadurch, daß jeder durch die vier Quantenzahlen gegebene Zustand nur durch ein Teilchen mit dem Spin 1/2 besetzt werden kann (vgl. 5.6.3.). Der Aufbau der Nukleonenschalen geht analog und getrennt für Protonen und Neutronen, setzt aber voraus, daß Spin und Bahndrehimpuls der Nukleonen stark gekoppelt sind. Man kann die Kopplungskonstante so wählen, daß bei den magischen Zahlen die Energieniveaus der Nukleonen zu stabileren Kernen führen. Kerne mit magischen Zahlen haben die experimentell gefundene Eigenschaft, schwerer als andere Kerne ein Nukleon abzugeben oder aufzunehmen. Damit ist es auch energetisch günstig und leicht möglich, wenn nur ein Nukleon zu einem Kern mit magischen Zahlen fehlt, dieses anzulagern. Das Schalenmodell der Kerne erklärt viele Effekte, so zum Beispiel die unsymmetrische Massenverteilung der beiden Bruchstücke bei der Kernspaltung; die Bruchstücke haben bevorzugt die Massenverteilung so, daß sich Kerne mit magischen Zahlen bilden. Zugleich zeigt gerade dieses Beispiel die Unzulänglichkeit unserer Vorstellungen und des Wissens über die Kernkräfte. Zur Kernspaltung ist das Tröpfchenmodell notwendig und zur Massenverteilung der Spaltprodukte das Schalenmodell sinnvoll.

5.2.4.3. Angeregte Kerne. Kerne können wie Atome Energie aufnehmen und in einen angeregten, mehr oder weniger stabilen Zustand übergehen. Für jeden Kern ergibt sich ein System von Anregungsstufen. Angeregte Zustände sind zum einen durch die Anregung eines oder einiger Nukleonen auf höhere Drehimpulswerte gekennzeichnet. Die Anregungsenergien können hierbei einige Megaelektronenvolt über dem Grundzustand liegen. Zum anderen kann der Kern Formschwingungen ausführen, z. B. zwischen gestrecktem und abgeplattetem Rotationsellipsoid um die Kugelform schwingen. Deformierte Kerne können unter Energieaufnahme rotieren, die Anregungsenergien sind geringer als für Einzelteilchenanregungen.

Isomere Kerne. Mitunter haben angeregte Kerne eine sehr geringe Wahrscheinlichkeit, in den Grundzustand überzugehen. Hat ein angeregter Kern eine Lebensdauer ($T_{1/2}$ = Halbwertszeit) von Stunden oder sogar Jahren (z. B. $^{193}_{77}Ir^m$, $T_{1/2} \approx 12$ Tage, $93\,Nb^m$, $T_{1/2} = 3{,}65$ Jahre), so existiert dieser angeregte Zustand mit eigenständigen Eigenschaften, etwa in den Zerfallsarten, durchaus verschieden vom Grundzustand, so daß man von einem Isomer bzw. einem isomeren Kern spricht. Er hat die gleiche Protonen- und Neutronenzahl wie der Kern im Grundzustand, aber andere Eigenschaften. Isomere Kerne werden mit einem hochgestellten „m" hinter dem Elementsymbol gekennzeichnet.

Beispiel: ^{234}Pa und $^{234}Pa^m$ unterliegen einem β^--Zerfall; das Isotop im Grundzustand mit einer Halbwertszeit von 6,7 Stunden, das zugehörige Isomer mit einer Halbwertszeit von 70,5 Sekunden.

Niveaubreiten, Halbwertszeiten. Unter der Breite eines Energieniveaus – *Niveaubreite* – versteht man seine Ausdehnung im Energiemaßstab. Sie wird als *Linienbreite* Γ der emittierten Strahlung in Höhe des halben Wertes der maximalen Intensität angegeben (Halbwertsbreite) (Abb. 5.2.4–1). Die Halbwertsbreite $T_{1/2}$ ist ein Maß für die *mittlere Lebendauer* τ des Anregungszustandes $T_{1/2} = \tau \cdot \ln 2$. Linienbreite und Lebensdauer eines Zustandes können nicht unabhängig voneinander beliebige Werte annehmen. Wirken keine anderen das Niveau verbreiternden Effekte,

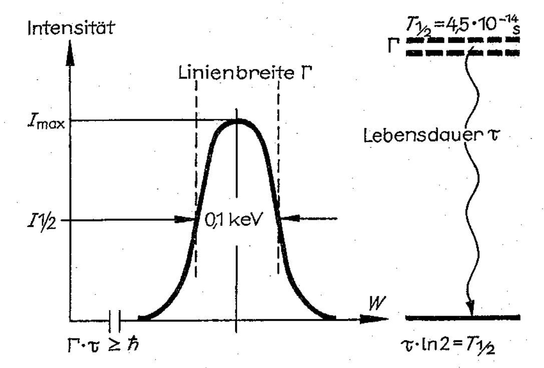


Abb. 5.2.4–1 Zusammenhang zwischen mittlerer Lebensdauer τ eines angeregten Zustandes und seiner Linienbreite Γ gemäß der Heisenbergschen Unbestimmtheitsbeziehung

ergibt sich die geringstmögliche Linienbreite nach der Heisenbergschen Unbestimmtheitsbeziehung, $\Gamma \cdot \tau \geqq \hbar$ zu $\Gamma \cdot T_{1/2} = 0{,}45 \cdot 10^{-15}$ [eVs].
Im Beispiel der Abb. 5.2.4–1 ist eine Linienbreite von 0,1 keV gemessen, woraus eine Halbwertszeit des angeregten Zustandes von $4{,}5 \cdot 10^{-14}$ s folgt. Bei der Messung ist zu beachten, daß die Niveaubreite des Endzustandes beim Übergang genügend scharf ist, da er sonst in die Linienbreite mit eingeht.
Möglichkeiten der Anregung von Kernen. Die Untersuchung der Kernanregungszustände ist für das Verständnis der Kerne so wesentlich wie die Analyse der Spektrallinien für den Atombau, für den Aufbau der Elektronenhülle. Kernanregungszustände ergeben sich in vielfältiger Weise. Bei der Kernspaltung entstehen auch Kerne in angeregten Zuständen, ebenso innerhalb der Zerfallsreihen. Kerne können auch direkt über die Absorption von γ-Quanten angeregt werden, wobei sowohl die Absorption als auch nachfolgende Emissionen analysiert werden können. Schnelle Neutronen und Protonen können dem Kern Energie übertragen, auch ohne selbst absorbiert zu werden (inelastische oder unelastische Nukleonenstreuung). Die Absorption von Nukleonen führt aufgrund der freiwerdenden Bindungsenergie zu hochangeregten Zuständen. Verlaufen Kernreaktionen über einen Zwischenkern (vgl. 5.3.2.4.), so sind dessen Anregungszustände meßbar. Die Zwischenkernbildung ist immer für Einschußenergien, die genau in Anregungszustände führen, wahrscheinlicher als für andere Energiewerte. Die angeregten Zustände äußern sich durch einen erhöhten Wirkungsquerschnitt in Abhängigkeit von der Einschußenergie.
5.2.4.4. Der Mößbauer-Effekt (rückstoßfreie Kernresonanzabsorption). In der Physik der Hülle ist die Fluoreszenz ein gutbekannter Effekt. Weiße Hemden können im UV-Licht einer Tanzbar bläulich strahlen, bestimmte Farben scheinen im Sonnenlicht selbstleuchtend zu sein, und „optische Aufheller“ werden Waschmitteln beigegeben. Eine mit Natriumdampf gefüllte Glaskugel strahlt ihr mildes, gelbes Licht gleichmäßig in alle Richtungen, wenn sie dem Lichtstrahl einer Na-Dampf-Lampe ausgesetzt wird. Die derart fluoreszierenden Substanzen absorbieren Lichtquanten einer ihnen eigenen Frequenz, und die dadurch angeregten Atome geben dann relativ rasch ihre Energie in Form der gleichen Lichtquanten wieder ab. Daß Atome die von ihnen emittierten Quanten auch wieder absorbieren können, erscheint selbstverständlich. Die Absorption erfolgt in Frequenzresonanz, die Frequenz des emittierten Lichtes ist gleich der des absorbierten Lichtes, man spricht von *Resonanzabsorption*. Mit Atomkernen müßte der analoge Effekt möglich sein. Die γ-Quanten emittierenden Niveaus müßten auch wieder mit den gleichen Quanten anregbar sein.
Die Verhältnisse in Kernen sind in zweierlei Hinsicht verschieden. Zum einen können die Linienbreiten wesentlich geringer sein, zum anderen sind die γ-Quantenenergien um Größenordnungen größer als die der Lichtquanten. Letzteres bewirkt, daß die auf das Atom übertragene Rückstoßenergie nicht mehr zu vernachlässigen ist; ersteres bewirkt, daß das mit verringerter Energie abgestrahlte Quant selbst die untere Grenze des sehr schmalen Niveaus nicht erreichen kann.
Ein γ-Quant der Frequenz ν_0 überträgt beim Verlassen des Kernes diesem einen Rückstoß von $p_R{}^2 = h\nu_0/c$, woraus sich die Rückstoßenergie zu $W_R = p_R{}^2/2M = (h\nu_0)^2/2Mc^2$ berechnet (M = Masse des Kernes). Für das in der Mößbauer-

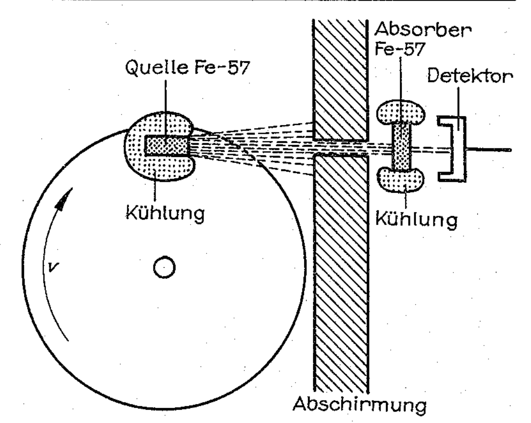


Abb. 5.2.4–2 Anordnung zur Messung der Kernresonanzabsorption

Spektroskopie oft verwendete γ-Quant eines Überganges des Fe-57-Isotopes von $h\nu_0 = 14{,}4$ keV beträgt die Rückstoßenergie $2 \cdot 10^{-3}$ eV. Um diesen Betrag ist die Energie des γ-Quantes geringer als der zugehörige Niveauzustand. Das ist nur $\approx 10^{-7}$ der Quantenenergie selbst, aber gegen die durch die Linienbreite gegebene zulässige Toleranz sehr viel. Die Linienbreite ist in diesem Falle mit $5 \cdot 10^{-9}$ eV besonders schmal, ein abgestrahltes Quant kann nicht ohne weiteres von der gleichen Atomsorte absorbiert werden. Das war schon in der Weise Anfang der 50er Jahre bekannt, und es gab eine intelligente Methode, den Energieverlust zu kompensieren. Um die Absorption zu erwirken, muß die Energie des emittierten Quantes um $2\,W_R$ vergrößert werden, da der gleiche Impuls auch bei der Absorption an das Atom übertragen wird. Eine Möglichkeit, die Energie des γ-Quantes geringfügig zu verändern, bietet der Doppler-Effekt. Wird die Strahlungsquelle mit der Geschwindigkeit v bewegt, so verschiebt sich die Frequenz bei Emission in Richtung der Bewegung zu $\nu = \nu_0(1 + v/c)$. Daraus resultiert ein Energieübertrag auf das Quant von $\Delta W = h\nu_0 \cdot v/c$. Für das oben betrachtete γ-Quant von 14,4 keV Energie beträgt der Wert $\approx 5 \cdot 10^{-5}$ eV für $v = 1$ m/s. Eine einfache mechanische Bewegung der Strahlungsquelle ist also geeignet, eine Absorption der von den gleichen Kernen emittierten γ-Quanten doch zu erreichen, die Resonanzabsorption zu erwirken. Das gelang auch mit der Anordnung gemäß Abb. 5.2.4–2, aber nur sehr unvollkommen. MÖSSBAUER kühlte 1958 Quelle und Absorber und erreichte damit den Durchbruch zu einer Meßtechnik, die in ihrer Präzision bis dahin unvorstellbar war. Durch die Tiefkühlung ist jede Bewegung der Atome in Quelle und Absorber unterbunden. Auch die in den Materialien angereicherten Fe-57-Atome sind ohne Möglichkeit, Eigenbewegungen auszuführen. Dadurch können sie allein nicht den Rückstoß aufnehmen, der Rückstoß wird vom gesamten Kristall übernommen. Da die Masse des Kristalles unendlich groß gegen die eines Atomes ist, wird die Rückstoßenergie praktisch gleich Null. Die derart *rückstoßfreie Kernresonanzabsorption* wird *Mößbauer-Effekt* genannt. Die Tiefkühlung von Quelle und Absorber verhindert auch jeglichen Energietransport über Phononen (Gitterschwingungen), so daß keine zusätzlichen Effekte die Linien verbreitern. Durch den nicht mehr zu kompensierenden Rückstoß führen schon sehr geringe Bewegungen der Quelle ($\approx 10^{-6}$ m/s) zur Verstimmung der

Resonanz. Die mit diesem Prinzip arbeitende Mößbauer-Spektroskopie kann extrem genau Energiedifferenzen messen. Manche Atomniveaus, wie das des Fe-57-Isotopes von 14,4 keV, sind so scharf, daß sie es gestatten, Frequenzänderungen von $\Delta\nu/\nu \approx 10^{-15}$ zu messen. Schönstes Beispiel dieser Spektroskopie ist der Nachweis der von EINSTEIN vorhergesagten Energieänderung von Strahlungsquanten unter dem Einfluß von Gravitationsfeldern. Abb. 5.2.4–3 zeigt den dafür verwendeten prinzipiellen Versuchsaufbau. Eine γ-Quelle strahlt senkrecht nach unten. Der Energiezuwachs ΔW ist durch das Produkt von Masse m_γ, Erdbeschleunigung g und Fallstrecke l gegeben, $\Delta W = m_\gamma g l$. Die Masse des γ-Quantes ist $m_\gamma = h\nu/c^2$. Für das γ-Quant des Fe-57 von $W_\gamma = 14{,}4$ keV errechnet sich der relative Energiezuwachs $\Delta W/W$ zu $1{,}1 \cdot 10^{-16}$ je 1 m Fallhöhe. In dem möglichst weit unten angebrachten Absorber kann mittels eines γ-Detektors die Resonanzabsorption nachgewiesen werden. Sie erreicht ihr Maximum, wenn der Energiezuwachs im Schwerefeld der Erde gerade durch die gegenläufige Bewegung der Quelle kompensiert wird. Aus der Bewegung der Quelle ergibt sich die durch den Doppler-Effekt erzeugte Frequenz- und damit Energieverschiebung.

Das Experiment ergab eindeutig, daß die bewegte Masse der Quanten elektromagnetischer Wellen in gleichem Maße wie Ruhmassen der Gravitationskraft unterliegen.

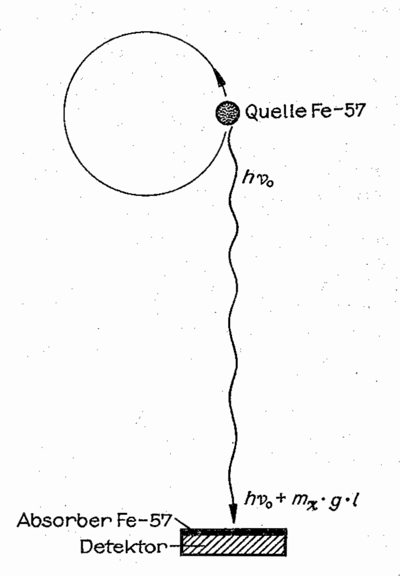


Abb. 5.2.4–3 Prinzip zur Messung des Energiezuwachses eines γ-Quants im Schwerefeld der Erde

5.3. Kernreaktionen

5.3.1. Durchgang elementarer Teilchen durch Materie

Beim Durchdringen von Gasen, Flüssigkeiten oder festen Stoffen – allgemein von Materie – verhalten sich die verschiedenen Arten der elementaren Teilchen verschieden. Zwar verlieren alle Teilchen Energie, werden dadurch abgebremst und aus ihrer Richtung ausgelenkt, aber die Art, Stärke und Häufigkeit der einzelnen Wechselwirkungsakte ist verschieden. Für alle Teilchen gilt ungefähr, daß sie um so durchdringender sind, je höher ihre Energie ist. Besonders durchdringend sind Neutronen und γ-Quanten. Nur 1/100 der Durchdringungsfähigkeit von γ-Strahlung haben Elektronen, die aber immer noch $\approx$ 100mal durchdringender sind als α-Teilchen.

5.3.1.1. Durchgang von γ-Strahlen durch Materie. Das *Absorptionsvermögen* energiereicher Photonen, die ab einigen 100 keV Energie als *γ-Quanten* bezeichnet werden, ist durch ein Exponentialgesetz beschreibbar. Die Intensität I_0 eines Strahles von γ-Quanten verringert sich nach Durchdringen einer Schichtdicke x auf I_x gemäß $I_x = I_0 \exp.\,(-\mu x)$. μ ist eine *Absorptionskoeffizient* genannte Materialkonstante.

Beispiel: μ (Blei, $W_\gamma = 1$ MeV) $= 8 \cdot 10^{-3}\,\mathrm{m}^{-1}$, woraus sich ergibt, daß die Intensität einer γ-Strahlung von 1 MeV Energie nach einer Schichtdicke Blei von $d = 8 \cdot 10^{-3}$ m (8 mm) auf $1/e = 0{,}37$, also nur um $\approx 2/3$ verringert wird.

Derartige exponentielle Zusammenhänge wie hier zwischen I_x und x gestatten die Einführung einer sehr praktischen Größe, die in diesem Falle Halbdicke ($X_{1/2}$) genannt wird (vergleiche auch Halbwertszeit 5.3.3.2.). Sie bezeichnet die Schichtdicke x, nach der die Intensität auf die Hälfte des Eingangswertes $I_x = I_0/2$ abgesunken ist, d.h. nach $2X_{1/2}$ auf 1/4 nach $3X_{1/2}$ auf 1/8, ... nach $10X_{1/2}$ auf $\approx 1/1000$ der ursprünglichen Intensität. Die Halbdicke ist aus dem Absorptionskoeffizienten zu berechnen, $X_{1/2} = 0{,}693/\mu$.

Drei Wechselwirkungseffekte sind wesentlich für die Abbremsung von γ-Quanten beim Durchgang durch Materie: Fotoeffekt, Compton-Streuung und Paarbildung.

Fotoeffekt ist die Ablösung von Elektronen aus dem Atom oder Atomverband durch Einwirkung elektromagnetischer Strahlung (äußerer Fotoeffekt). Das γ-Quant wird absorbiert und seine Energie als kinetische Energie auf ein Elektron übertragen. Die Absorption eines γ-Quants oder Photons ist nur an gebundene Elektronen möglich, da sonst nicht gleichzeitig Energie und Impuls der elektromagnetischen Strahlung vollständig abgegeben werden können. Allerdings ist die Energieübertragung an das Atom sehr gering, allgemein vernachlässigbar. Es gilt in guter Näherung, daß die kinetische Energie des Elektrons W_{kin} gleich der Energie des γ-Quantes W_γ minus der Ablösearbeit W_B vom Atom bzw. Atomverband ist, $W_{\text{kin}} = W_\gamma - W_B$.

Compton-Streuung ist ein Effekt der Teilcheneigenschaft der elektromagnetischen Strahlung. Das γ-Quant stößt in Art eines Billard-Stoßes ein Elektron an und überträgt ihm kinetische Energie. Das γ-Quant wird dabei aus seiner Richtung ausgelenkt und der Energieverlust zeigt sich in einer Vergrößerung der Wellenlänge. Die Streuung genügt der Energie- und Impulserhaltung.

Paarbildung bedeutet die Vernichtung eines γ-Quants unter Bildung eines Elektrons und eines Positrons. Analog dem Fotoeffekt ist die Anwesenheit eines dritten Stoßpartners erforderlich. Die Elektron-Positron-Paarbildung erfordert nach der Einsteinschen Masse-Energieäquivalenz $W = Mc^2$, entsprechend der Ruhmasse des Teilchenpaares ($M_{\text{Elektron}} = M_{\text{Positron}} = 0{,}511\,\mathrm{MeV}/c^2$), eine Mindestenergie von 1,02 MeV. Erst ab dieser Energie für ein γ-Quant kommt es zur Paarbildung.

Die drei Effekte Foto-, Compton- und Paarbildungseffekt erklären im wesentlichen den Energieverlust, die Absorption von γ-Quanten beim Durchgang durch Materie. In Abb. 5.3.1–1 sind die Effekte einzeln in Abhängigkeit von der γ-Quantenenergie dargestellt.

Aufgetragen ist der Wirkungsquerschnitt – ein Maß für die Stärke des Auftretens der Effekte. Durch Addition der einzelnen Effekte ergibt sich der Gesamtwirkungsquerschnitt

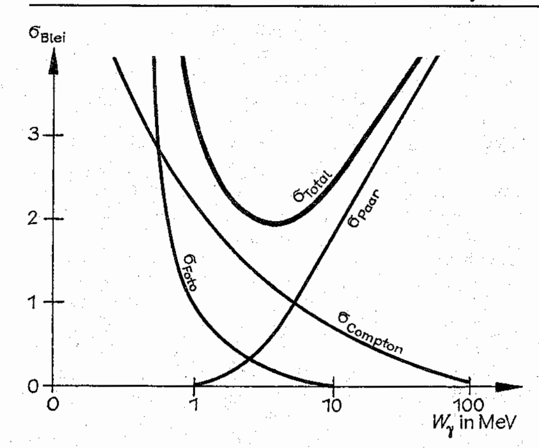


Abb. 5.3.1–1 Schematische Darstellung der Wirkungsquerschnitte für Fotoeffekt σ_{Foto}, Compton-Effekt σ_{Compton}, Paarbildung σ_{Paar} und der totale Wirkungsquerschnitt $\sigma_{\text{total}} = \sigma_{\text{Foto}} + \sigma_{\text{Compton}} + \sigma_{\text{Paar}}$ für Blei in relativen Einheiten

σ_{total}. Der Fotoeffekt, dargestellt durch σ_{Foto}, überwiegt bei niederen Quantenenergien und in Abhängigkeit von der Ordnungszahl Z der Elemente besonders stark bei schweren Elementen: $\sigma_{\text{Foto}} \sim Z^5/W_\gamma^{+3,5}$. Der Wirkungsquerschnitt für die Compton-Streuung $\sigma_{\text{Compton}} \sim Z/W_\gamma$ ist, da durch die Zahl der vorhandenen Elektronen bestimmt, proportional der Ordnungszahl Z. Daß die Paarbildung bei hohen Energien den Energieverlust bestimmt, erkennt man auch am zugehörigen Wirkungsquerschnitt: $\sigma_{\text{Paar}} \sim Z^2 \cdot \ln W_\gamma$, der mit zunehmender Energie ansteigt.

5.3.1.2. Durchgang geladener Teilchen durch Materie. Ionisationsverluste. Die Wechselwirkung geladener Teilchen in Gasen, Flüssigkeiten und festen Körpern erfolgt überwiegend über Coulomb-Kräfte – es wechselwirken die Ladungen. Geladene Teilchen stoßen längs ihres Weges Elektronen aus ihren Ruhelagen, bilden Elektron-Ion-Paare, d. h. sie ionisieren. Bei gleicher kinetischer Energie ionisieren Protonen stärker als Elektronen und α-Teilchen stärker als Protonen. Elektronen haben den kleinsten Wert für die *spezifische Ionisation*, unter der man die je Millimeter Weglänge gebildete Zahl der Elektron-Ion-Paare versteht. Sie beträgt für α-Teilchen in Luft ≈ 2000, für Elektronen hingegen nur $\approx 6{,}5$ Elektron-Ion-Paare je Millimeter Weglänge.

Strahlungsbremsung. Auf diesen Effekt beruht die Erzeugung von Röntgen-Strahlung (vgl. 5.6.5.). Allgemein können alle geladenen Teilchen in dem starken Coulomb-Feld der Atomkerne Bremsverlust unter Aussendung von γ-Quanten erleiden. Diese Strahlungsbremsung ist aber umgekehrt proportional dem Quadrat der Masse der geladenen Teilchen, so daß es überwiegend ein Effekt der mit Materie wechselwirkenden Elektronen ist. In großen Beschleunigern und Speicherringen, wo die geladenen Teilchen auf gekrümmten Bahnen laufen, bewirkt die Zentrifugalbeschleunigung eine Energieabstrahlung, die als Quelle harter Röntgen-Strahlen vielfältig angewendet wird (vgl. 5.4.3.4.).

Čerenkov-Strahlung entsteht durch einen Effekt, der wegen seiner Kleinheit allgemein bei der Abbremsung von Teilchen nicht berücksichtigt wird. Bewegt sich ein Teilchen in einem Medium (des Brechungsindexes n) schneller als Licht, also mit einer Geschwindigkeit $v > c/n$, wobei c die Vakuumlichtgeschwindigkeit ist, strahlt es aus einem sich hinter ihm öffnenden Kegel Licht ab – ganz analog der Schallschleppe hinter mit Überschallgeschwindigkeit bewegten Flugkörpern. Es ergibt sich der gleiche Zusammenhang zwischen Teilchengeschwindigkeit v, Lichtgeschwindigkeit c und dem Machschen Winkel α (halber Öffnungswinkel des Licht- bzw. Schallkegels), $\cos\alpha = c/(n \cdot v)$. Die Abhängigkeit der Richtung des abgestrahlten Lichtes von der Geschwindigkeit des Teilchens kann zu seiner Energiebestimmung genutzt werden (Čerenkov-Zähler, vgl. 5.4.4.1.). In Kernreaktoren mit Wasser als Moderatorsubstanz kann man in dem die Brennstäbe umgebendem Wasser diese Strahlung als bläuliches Leuchten beobachten.

5.3.1.3. Durchgang von Neutronen durch Materie. Durch die fehlende elektrische Ladung der Neutronen ist ihre Wechselwirkung mit Materie wesentlich geringer und von anderer Art als bei geladenen Teilchen. Schnelle Neutronen, wie sie z. B. bei der Tritium-Deuterium-Fusion mit 14 MeV kinetischer Energie entstehen, haben in Luft eine Reichweite von knapp einem Kilometer. Bei ihrer Abbremsung kommen im wesentlichen nur zwei Arten der Wechselwirkung in Betracht: elastische Streuung und Absorption.

Elastische Streuung ist im Gegensatz zu den inelastischen Reaktionen, wo die innere Energie des Kernes, Atomes oder Atomverbandes geändert wird, nur eine Übertragung von kinetischer Energie. Die elastischen Stöße folgen den Gesetzen der klassischen Mechanik, wie etwa der Stoß von Billardkugeln. Ein technisch sehr wichtiger Prozeß in Kernreaktoren ist die Abbremsung von schnellen Neutronen (Moderation). Am effektivsten geschieht das bei Stößen mit den gleichschweren Protonen, den Wasserstoffatomen. Um Neutronen von 1-MeV-Energie auf thermische Energien (0,025 eV entsprechend ≈ 300 K) abzubremsen, genügen ≈ 17 Stöße. Nicht mehr so effektiv ist die Energieabgabe bei Stößen mit schwereren Teilchen. Tab. 5.3.1–2 gibt die für die Moderation von Reaktorneutronen notwendige Zahl von Stößen n für verschiedene Elemente der Massenzahl A an. Je leichter das Element, desto schneller wird das Neutron abgebremst.

Tab. 5.3.1–2 Zahl der zur Abbremsung (Moderation) von schnellen Neutronen mit einer Energie von 1 MeV auf 0,025 eV notwendigen Stöße n für verschiedene Elemente der Massenzahl A

	A	n
H	1	17,5
D	2	23,5
He	4	40
Be	9	81
C	12	108
Pb	207	1770

Die Absorption von Neutronen ist einerseits ein erwünschter Effekt beim Schutz vor Strahlenschäden, andererseits ein störender Effekt bei der Moderation im Reaktor, wo man die moderierten (die langsamen) Neutronen zur Kernspaltung benötigt.

Die Absorption erfolgt über Kernreaktionen und ist für verschiedene Materialien sehr unterschiedlich. In Tab. 5.3.1–3

sind die Einfachquerschnitte für thermatisierte Neutronen für verschiedene Materialien angegeben. Ein hervorragender Absorber, ein regelrechtes Reaktorgift, ist das leicht verfügbare Cadmium. Ein schlechter Absorber, an dem sich praktisch keine Neutronen ablagern, ist Deuterium, weshalb in Kernreaktoren zur Moderation auch schweres Wasser (D_2O) verwendet wird.

Tab. 5.3.1–3 Wirkungsquerschnitte für die Absorption thermischer Neutronen in barn (10^{-28} m²)

H	0,33
D	0,0006
B	755
Fe	2,6
Cd	25000
Gd	46000
U-235	107

5.3.1.4. Energie-Reichweite-Beziehungen. Betrachtet man die Summe aller Wechselwirkungen eines Teilchens mit Materie, so resultiert daraus der Energieverlust $-\mathrm{d}W/\mathrm{d}x$, für ein Weglängenelement dx. Dieser *differentielle Energieverlust* ist von der Teilchenart, seiner Energie und der Art der durchdrungenen Materie abhängig. Ein Teilchen verliert Energie, bis es nach einer bestimmten Wegstrecke zur Ruhe kommt. Die gesamte Wegstrecke, die Integration über alle Energieverluste bis zur Teilchenenergie 0, ergibt

die Reichweite R: $-\frac{\mathrm{d}W}{\mathrm{d}x} = \frac{\mathrm{d}W}{\mathrm{d}R}$; $R = \int_{W_{\mathrm{kin}}}^{0} -\left(\frac{\mathrm{d}W}{\mathrm{d}x}\right)^{-1} \mathrm{d}W.$

Die Theorie liefert Ausdrücke für die differentiellen Energieverluste $-\mathrm{d}W/\mathrm{d}x$, die die schon länger bekannten empirischen Energie-Reichweite-Beziehungen begründen. Die Beziehungen zwischen Energie und Reichweite ergeben meist in großen Bereichen Geraden, wenn man logarithmische Darstellungen wählt, also Beziehungen der Form $W = a \cdot R^m$.

Beispielsweise gilt für α-Teilchen in Luft die Beziehung $R_\alpha = 0{,}00318 \cdot W_\alpha^{1,5}$, R_α ist in Meter, W_α in Megaelektronenvolt angegeben. Die Gleichung gilt zwischen einigen 10^{-2} und einigen 10 m Spurlänge in Luft.

Abb. 5.3.1–4 ist eine Darstellung der Reichweiten von Elektronen, Protonen und α-Teilchen in Luft und Blei in Abhängigkeit von der Energie der Teilchen. Der Abbildung entnimmt man, daß α-Teilchen aus radioaktiven Präparaten (mit maximal 10 MeV) durch eine Bleischicht von 0,01 mm Dicke abgeschirmt werden, für Protonen der gleichen Energie schon mehr als 0,1 mm und für Elektronen bereits einige Millimeter notwendig sind. γ-Quanten der gleichen Energie würden erst durch einige Zentimeter Blei auf 1% ihrer ursprünglichen Intensität zu verringern sein.

Die Reichweite wird oft nicht in der Dimension Meter, sondern als Flächengewicht (mit der Dichte ϱ multipliziert) in kg/m² angegeben. Das hat den Vorteil, etwa für die Abb. 5.3.1–4, nur eine Skala zu benötigen.

Über diese Beziehung läßt sich, wenn man annimmt, die Teilchenart zu kennen, aus der Reichweite die Energie bestimmen. Hat man derart die Anfangsenergie eines Teilchens beim Eindringen in ein bestimmtes Material ermittelt, so läßt sich aus dem spezifischen Energieverlust dort – also $-\mathrm{d}W/\mathrm{d}x$ zu Anfang der Wegstrecke –, da auch dieser von der Teilchenart abhängt, die Teilchenart bestätigen.

Wegen der sehr verschiedenen Reichweiten der verschiedenen Teilchenarten bei gleicher Energie ist auch der spezifische Energieverlust sehr unterschiedlich. Etwa den gleichen spezifischen Energieverlust – dE/dx – haben z.B. Elektronen von 10 keV, Protonen von 30 MeV und α-Teilchen von 500 MeV Energie.

5.3.2. Allgemeines zu Kernreaktionen

5.3.2.1. Wirkungsquerschnitt, freie Weglängen. Zur Beschreibung der Häufigkeit des Auftretens bestimmter Reaktionen – etwa des Neutroneneinfanges – eigenet sich der *Wirkungsquerschnitt.* Ein Strom von Teilchen dN/dt durchdringe ein Material der Fläche 1 m² und der Dicke d. Hinter diesem Material sei dN_i/dt die Anzahl der Teilchen je Zeit-

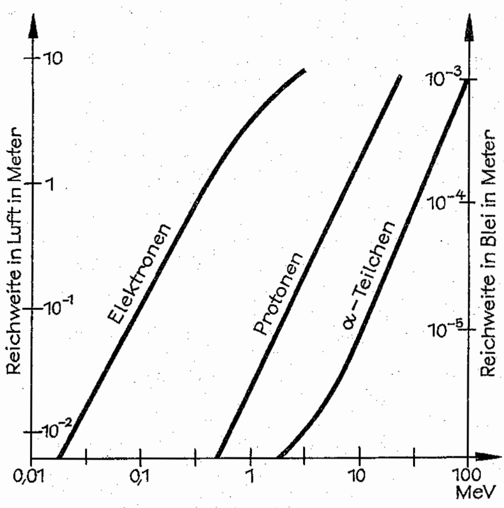


Abb. 5.3.1–4 Energie-Reichweite-Beziehungen für Elektronen, Protonen und α-Teilchen in Luft und Blei.

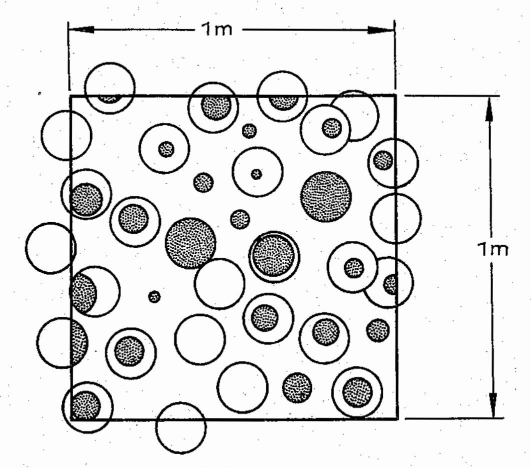


Abb. 5.3.2–1 Geometrischer Wirkungsquerschnitt als Fläche der „verschnittenen" Atome auf 1 m² (punktiert)

einheit, die in bestimmter Weise (Index i) wechselwirkten.

Dieser Sachverhalt wird durch $dN_i/dt = \sigma dN/dt \cdot n \cdot d$ beschrieben, wobei $n\,[m^{-3}]$ die zur Auslösung der Wechselwirkung im Material vorhandene Teilchendichte darstellt. σ ist ein Wirkungsquerschnitt genannter Proportionalitätsfaktor der Gleichung mit der Dimension m^2. Anschaulich ist er die „Trefferfläche", gebildet aus dem Quotienten der Zahl der Reaktionen zur Zahl der auf die Einheitsfläche unternommenen Versuche. Durchschneidet man in Gedanken ein Material, so bilden die auf der Schnittfläche von $1\,m^2$ sichtbaren Schnittflächen der Atome den „geometrischen" Wirkungsquerschnitt (Abb. 5.3.2–1).

Wirkungsquerschnitte für Kernreaktionen liegen in den Größenordnungen um $10^{-28}\,m^2 = 1$ barn, einer in der Kernphysik gebräuchlichen Einheit für den Wirkungsquerschnitt. Wirkungsquerschnitte sind für alle Wechselwirkungen definierbar. Es gibt z. B. Streuquerschnitte, Spaltquerschnitte, Querschnitte für bestimmte – z. B. n, α-Reaktionen. Interessant ist, daß Wirkungsquerschnitte nichts mit der Größe der Atome oder Atomkerne zu tun haben müssen. So sind z. B. die Neutroneneinfangquerschnitte beim Cadmium einige tausendmal größer und beim Deuterium drei Größenordnungen kleiner als der geometrische Wirkungsquerschnitt. Beim Deuterium treffen auch Neutronen die Deuterium-Atome, ohne sich anlagern zu können, beim Cadmium werden die Neutronen aus einem großen Umkreis regelrecht eingefangen (vgl. Tab. 5.3.1–3).

Mittlere freie Weglänge. Die Strecke, die ein Teilchen im Mittel zwischen zwei Wechselwirkungen zurücklegt, wird *mittlere freie Weglänge* genannt und oft mit λ bezeichnet. Sie ist genau so speziell oder allgemein wie Wirkungsquerschnitte für alle Reaktionen angebbar und zusammen mit der in einem Material vorhandenen Teilchendichte n mit dem Wirkungsquerschnitt verknüpft zu $\lambda = (n \cdot \delta)^{-1}$.

Je größer die Dichte eines Materials und je größer der Wirkungsquerschnitt, umso geringer ist die mittlere freie Weglänge.

5.3.2.2. Arten von Kernreaktionen. Unter *Kernreaktionen* versteht man Wechselwirkungen mit Kernen, die zu den nachfolgenden Arten gehören.

Austauschreaktionen. Ein Teilchen a wechselwirkt mit dem Kern A, wobei als Endprodukt ein Kern B und ein Teilchen b entsteht, allgemein geschrieben: A (a, b) B. Als Beispiel sei das historische Experiment der ersten künstlichen Kernumwandlung angeführt, $^{14}_{7}N\ (\alpha, p)\ ^{17}_{8}O$, ausführlich: $^{14}_{7}N + \alpha \rightarrow {}^{17}_{8}O + p$.

Einfangreaktionen. Ein Teilchen a wechselwirkt mit dem Kern A, im Ergebnis entsteht ein Kern B und ein oder mehrere γ-Quanten, allgemein geschrieben, A(a, γ)B.

Als Beispiel für diese Verschmelzungsreaktionen (Fusionen) sei hier die Tritium-Erzeugungsreaktion $^{2}_{1}D\,(n, \gamma)\,^{3}_{1}T$ angeführt, ausführlich geschrieben, $^{2}_{1}D + n \rightarrow {}^{3}_{1}T + \gamma$.

Kernfotoeffekt. Durch Einstrahlung eines γ-Quantes in den Kern wird ein Kernteilchen b emittiert, der Kern B bleibt zurück, A (γ, b) B. Beispiel: $^{31}P\,(\gamma, n)\,^{30}P$. Diese Neutronenerzeugungsreaktion entspricht der Reaktionsgleichung: $^{31}p + \gamma \rightarrow {}^{30}p + n$.

Unelastische Streuung. Hierbei wird ein Teilchen a an einem Kern X gestreut. Dabei nimmt der Kern Energie auf und bleibt in einem angeregten Zustand zurück. X (a, a) X, z. B. in der inelastischen α-Teilchen-Streuung an Aluminium, $Al + \alpha \rightarrow Al^* + \alpha$.

Kernspaltung. In irgendeiner Weise, beispielsweise durch die Einstrahlung von Neutronen, wird einem Kern Energie zugeführt, worauf dieser in Schwingung gerät, sich überdehnt und in zwei Teile zerfällt. Beispiel sei hierfür die Spaltung des mit langsamen Neutronen leicht spaltbaren Uraniumisotopes U-235. Es entstehen zwei Bruchstücke und im Mittel zwei Neutronen, beispielsweise $^{235}_{92}U + n \rightarrow {}^{97}_{38}Sr + {}^{136}_{54}Xe + 2n$.

Kernzertrümmerung. Reaktionen, die durch Teilchen höchster Energie ausgelöst werden, etwa durch Höhenstrahlung, können ein regelrechtes Zerplatzen getroffener Kerne auslösen. Dabei entstehen viele Bruchstücke, deren Gesamtmasse bedingt durch die hohe Energie des einkommenden Teilchens, die ursprüngliche Masse wesentlich überschreiten kann.

5.3.2.3. Notwendige Energien zur Auslösung von Kernreaktionen. Um bis in die Reichweite der Kernkräfte vordringen zu können, sind je nach Art der reagierenden Teilchen verschiedene Energiebilanzen zu beachten. Es existieren prinzipielle Unterschiede für Reaktionen, die durch γ-Quanten, Neutronen oder geladene Teilchen ausgelöst werden. Tab. 5.3.2–2 gibt in einer Übersicht die Bedingungen zur Auslösung von Kernreaktionen durch verschiedene elementare Teilchen an.

Tab. 5.3.2–2 Bedingungen zur Auslösung von Kernreaktionen

Einschuß mit	Reaktion	Bedingung
γ-Quanten	A(γ, b)B	$W_\gamma \geqq W_B(b)$
Neutronen	A(n, b)B	$W_n \gtrsim 0$
p-, α-Teilchen	A(α, b)B	$W_\alpha \gtrsim W_{Coulomb}$

Will man mit γ-Quanten ein Teilchen, z. B. ein Neutron, aus dem Atomkern lösen, so ist mit der Energie des γ-Quantes mindestens die Energie, mit der das Neutron im Kern gebunden ist $[W_B(n)]$, aufzubringen. Derart sind γ-quanteninduzierte Kernreaktionen eine ideale Möglichkeit, Bindungsenergien zu bestimmen. So ist die Bindungsenergie des Neutrons gleich der γ-Quanten-Energie, ab der die Kernreaktion, d. h. die Emission eines Neutrons, einsetzt.

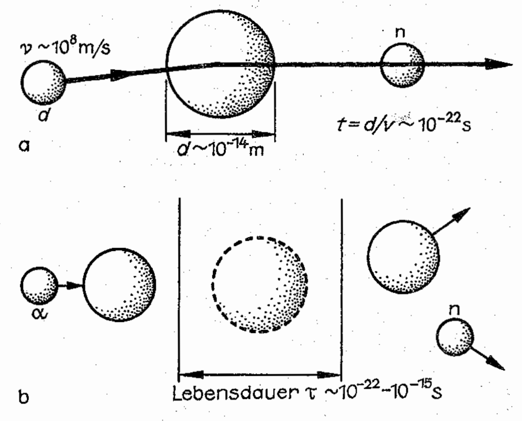


Abb. 5.3.2–3 Arten von Kernreaktionen, (a) Direktreaktion (Reaktionszeit ~ Durchflugszeit t), (b) Zwischenkernreaktion

Bei neutroneninduzierten Reaktionen wird durch Anlagerung des Neutrons an den Kern die zugehörige Bindungsenergie frei. Sie reicht oft aus, die Emission eines (z. B. α-Teilchens) zu bewirken. So erfolgt der Nachweis langsamer Neutronen z. B. durch α-Teilchen über die Reaktion $^{10}B(n, \alpha)^{7}Li$; $W_\alpha \geqq 2{,}78$ MeV.

Sollen mit geladenen Teilchen (p-, d-, α-Teilchen) Kernreaktionen ausgelöst werden, so ist die elektrostatische Abstoßung zwischen dem positiv geladenen Geschoßteilchen und dem ebenfalls positiv geladenen Kern zu überwinden. Die notwendigen Geschoßenergien W liegen in der Größenordnung von Megaelektronenvolt. Für α-Teilchen ergibt sich $W_\alpha \geqq 1{,}1 \cdot A^{2/3}$ [MeV], wobei für A die Massenzahl des zu beschießenden Kernes einzusetzen ist.

Elektronen können nicht über Kernkräfte wechselwirken; sie können ebensowenig auf diese reagieren, wie ungeladene Teilchen auf Ladungen.

5.3.2.4. Direktreaktionen, Zwischenkernreaktionen. Bei Kernreaktionen unterscheidet man zwischen 1. prompt ablaufenden Prozessen, die praktisch ohne Zeitverzögerung, sozusagen im „Durchfliegen" des Kernes, stattfinden und deshalb *Direktreaktionen* genannt werden, und 2. Reaktionen, bei denen ein Geschoß in den Kern einfliegt, mit diesem verschmilzt, einen angeregten Zustand bildet und nach einer gewissen Zeit zerfällt. Der gebildete Zwischenkern, auch Compound-Kern genannt, existiert aber nicht lange genug, um als eigenständiges Element akzeptiert werden zu können. Abb. 5.3.2–3 skizziert diese beiden Reaktionstypen.

5.3.3. Stabilität und Zerfall der Atome

5.3.3.1. Stabilitätsgrenzen und Zerfallsarten. In einem Diagramm, in dem die Zahl der Neutronen über die der Protonen aufgetragen ist, in dem man alle stabilen, instabilen und auch nicht existierenden Nuklide (Atomkerne, gebildet aus den Nukleonen, Protonen und Neutronen) eintragen kann, mag es den unbefangenen Betrachter verwundern, daß nur so wenige stabile Atome existieren. Man kann eine Kurve bilden, auf der bzw. dicht bei der die stabilen Atome angeordnet sind. Es sind nur ≈ 270,88 natürliche Elemente und ihre stabilen Isotope. Hinzu kommen noch ≈ 50 instabile Nuklide als Zerfallsprodukte der in der Natur vorkommenden radioaktiven Elemente. Künstlich erzeugt hat man nun bald die Hälfte der, von theoretischen Betrachtungen her, überhaupt möglichen 4000 Kerne.

In Abb. 5.3.3–1 ist in einem P-N-Diagramm die Linie, auf oder dicht bei der die stabilen Nuklide zu finden sind, eingetragen. Bei leichten Elementen ist die Zahl der Neutronen dort etwa gleich der der Protonen. Bei schweren Elementen gehören mehr Neutronen als Protonen zu den stabilen Kernen. Ob ein Kern stabil ist oder nicht, ergibt sich aus der Bindungsenergie (vgl. 5.2.3.). Je größer die Bindungsenergie der Nukleonen im Kern, umso stabiler ist der Kern, das Atom. Daß überhaupt so viele verschiedene Atome existieren, liegt nicht daran, daß die Nukleonen in all diesen Kernen gleichstarke Bindung haben, sondern daran, daß bei nur geringen Bindungsenergieunterschieden die möglichen Umwandlungen sehr unwahrscheinlich sind. Die Umwandlungen passieren um so eher, je größer der Gewinn an Bindungsenergien dabei ist, also je fester die Nukleonen nachher gebunden sind.

Das schwerste stabile Element ist Blei. Mehr als 82 Protonen sind aufgrund ihrer elektrostatischen Abstoßungskräfte

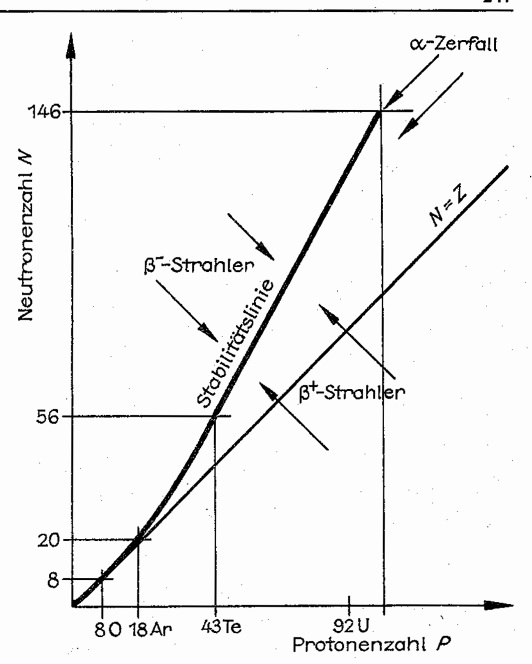


Abb. 5.3.3–1 Darstellung der Stabilitätslinie im N-P-Diagramm. Auf oder dicht bei ihr befinden sich die stabilen Elemente, etwas abseits die instabilen, weiter entfernt sind keine, auch keine kurzlebigen, Elemente existent

nicht auf Dauer zusammen zu halten. Aber auch die Atome unterhalb und oberhalb der Stabilitätslinie in Abb. 5.3.3–1, die mit Protonen- oder Neutronenüberschuß, sind nicht stabil. Instabile Nuklide wandeln sich je nach Art der Instabilität in verschiedener Weise in stabile Elemente um.

Zerfallsarten werden durch die Art der beim Zerfall auftretenden Strahlung charakterisiert.

α-Zerfall, ein Zerfall, eine Umwandlung in stabilere Elemente, unter Aussendung von α-Teilchen ist charakteristisch für zu schwere Elemente. In Abb. 5.3.3–1 ist er eine Umwandlung in Richtung der Stabilitätslinie parallel zur Geraden $N = Z$.

In einem α-Teilchen sind die Protonen und Neutronen stärker gebunden als in schweren Kernen. Dadurch kann von diesen ein α-Teilchen mit Energiegewinn für die restlichen Nukleonen emittiert werden. Hier muß aber der Energiegewinn so groß sein, daß das α-Teilchen auch noch genügend kinetische Energie mitbekommt, um mit ausreichender Wahrscheinlichkeit durch den Coulomb-Wall tunneln zu können.

β-Zerfall. Wenn instabile Elemente mit zu vielen Neutronen oder Protonen existieren, so sind selbst die überzähligen Teilchen so fest an den Kern gebunden, daß sie diesen nicht verlassen können. Damit ein Kern ein Proton oder Neutron emittieren kann, muß ihm die Bindungsenergie für das Nukleon zugeführt werden. Die Wandlung in stabilere Kerne geht ohne Änderung der Massenzahl, also zwischen isobaren Kernen (vgl. 5.2.3.4.) vor sich. Bei Neutronenüberschuß wandelt sich ein Neutron unter Aussendung eines Elektrons in ein Proton um (β^--Strahler), bei Protonenüberschuß entsteht unter Emission eines Positrons aus

einem Proton ein Neutron (β^+-Strahler). Bei beiden Zerfällen entstehen noch Neutrinos. Beim β^+-Zerfall $p \rightarrow n + e^+ + \nu$ das Neutrino ν, beim β^--Zerfall $n \rightarrow p + e^- + \bar{\nu}$ das Antineutrino $\bar{\nu}$.
Während die Umwandlung eines Neutrons in Proton, Elektron und Antineutrino exotherm, unter Abgabe von Energie an die Umwelt erfolgt, ist der β^+-Zerfall eine endotherme Reaktion. Die Massenbilanz weist aus, daß die Energie zur Bildung von zwei Elektronenmassen noch eingebracht werden muß. Steht diese Energie von 1,02 MeV nicht zur Verfügung, gibt es eine andere Möglichkeit der Wandlung, K-Einfang genannt. Der Kern kann sich ein Elektron der Hülle (K-Schale) einfangen und so auch aus einem Proton ein Neutron bilden, $p + e^- \rightarrow n + \nu$.
Eine weitere Art des Zerfalls instabiler Elemente ist die Kernspaltung. Schwere Kerne können selbst ohne äußere Einwirkungen (spontan) in zwei etwa gleichschwere Teile zerfallen, wodurch sich zwei mittelschwere Elemente bilden.
Die Grenze zwischen stabilen und instabilen Elementen ist mit letzter Konsequenz nicht genau feststellbar. Das Bismutisotop mit der Massenzahl 209 ($^{209}_{83}Bi$) ist α-aktiv. Aber die Zerfallswahrscheinlichkeit ist sehr gering. Erst in 10^{18} Jahren verringert sich eine vorgegebene Menge auf die Hälfte. Dadurch kann man das seltene Ereignis eines Zerfalles nur schwer beobachten. Vor einigen Jahren war durch diese Halbwertszeit von 10^{18} Jahren die Grenze der Meßbarkeit gegeben. Heute kann man mehr, aber wie sollte man die Grenze zur Instabilität bestimmen, wo soll man die Grenze legen? Derzeit sucht man die Halbwertszeit des Protons von $\approx 10^{31}$ Jahren zu finden. Gemessen am Alter der Erde von 10^9 Jahren ist das Proton bestimmt stabil.

5.3.3.2. Natürliche Radioaktivität. Unter *Radioaktivität* versteht man die Eigenschaft instabiler Kerne, beim Zerfall Strahlung auszusenden. Das können α-Teilchen, Elektronen, Positronen oder auch γ-Quanten sein. Die natürliche Radioaktivität entsteht unabhängig von menschlicher Einflußnahme.
Der radioaktive Zerfall läßt sich durch das Zerfallsgesetz beschreiben. Die zeitliche Folge der Zerfälle ist das Musterbeispiel einer Zufallsfolge. Auf ein einzelnes Atom bezogen, kann man nicht sagen, wann es zerfällt. Das Atom altert nicht. Und wenn es zerfällt, dann unabhängig von allen anderen.

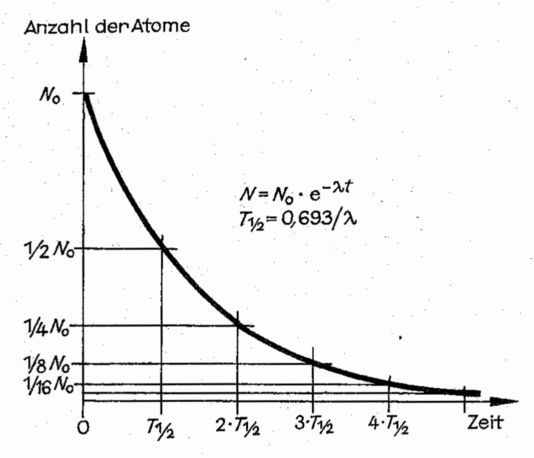


Abb. 5.3.3–2 Zeitliche Abnahme der Zahl der Atome eines radioaktiven Elementes (Zerfallsgesetz). Die Konstruktion der Kurve erfolgte mittels der Halbwertszeit $T_{1/2}$

Das Zerfallsgesetz hat zur Voraussetzung, daß die Zahl der Zerfälle dN/dt proportional der Zahl der Atome N ist, $dn/dt \sim N$. Daraus leitet sich ab, daß die Zahl der ursprünglichen Atome exponentiell mit der Zeit abnimmt. Die Zahl der zum Zeitpunkt t noch vorhandenen Atome $N(t)$ ist gleich der Zahl der Atome N_0 zum Zeitpunkt $t = 0$ multipliziert mit $\exp(-\lambda t)$, $N(t) = N_0 e^{-\lambda t}$ *(Zerfallsgesetz)*, λ ist die atomspezifische *Zerfallskonstante*.
Eine überaus praktische Größe ist die *Halbwertszeit* $T_{1/2}$, die Zeit, nach der die Substanz auf die Hälfte ($N_0/2$) der ursprünglichen Menge zerfallen ist. Sie ist mit der Zerfallskonstanten verknüpft, $T_{1/2} = \ln 2/\lambda = 0{,}693/\lambda$.
Abb. 5.3.3–2 zeigt, wie sich mit der Halbwertszeit die exponentielle Zerfallskurve konstruieren läßt. Nach einer Halbwertszeit existieren noch $1/2\,N_0$ Atome, nach der zweiten Halbwertszeit noch von $1/2\,N_0$ die Hälfte, also $1/4\,N_0$. Diese Menge ($1/4\,N_0$) betrachtet man nun wieder als Startmenge und nach der Zeit $T_{1/2}$ ist davon dann nur noch die Hälfte da usw.
Halbwertszeit bzw. Zerfallskonstante können für die einzelnen radioaktiven Elemente sehr verschieden sein. Tab. 5.3.3–3 zeigt dies. Die Tabelle enthält beispielhaft fünf radioaktive, α-Teilchen ausstrahlende Elemente mit Halbwertszeiten von $3 \cdot 10^{-7}$ s bis zu $4{,}5 \cdot 10^9$ Jahren. Der Tab. entnimmt man noch, daß die Energie der α-Teilchen um so größer wird, je geringer die Halbwertszeiten der Atome sind. Die Halbwertszeiten bzw. Zerfallskonstanten lassen sich berechnen. Die Theorie des α-Zerfalls beruht auf dem quantenmechanischen Tunneleffekt (vgl. 4.4.1.).

Tab. 5.3.3–3 Halbwertszeiten und Energien verschiedener α-aktiver Elemente

Element	$T_{1/2}$	W_α in MeV
^{212}Po	0,3 µs	8,78
^{212}Bi	60,5 m	6,09
^{222}Rn	3,8 d	5,49
^{226}Ra	1620 a	4,78
^{238}U	$4{,}5 \cdot 10^9$ a	4,19

Zerfallsreihen. Zerfällt ein Atom, so kann ein stabiles Element entstehen. Wenn nicht, geht der Zerfall weiter, es entstehen Zerfallsreihen. Die überwiegende Mehrzahl der 45 natürlichen radioaktiven Elemente gehört zu vier Zerfallsreihen. Als Beispiel sei die Uranium-Radium-Reihe angeführt. Die Reihe beginnt mit dem α-Zerfall des Uranium-238, erreicht nach vier weiteren Zerfällen das Radium, welches auch α-aktiv ist und weiter zerfällt. Die Zerfallsreihe endet bei dem stabilen Bleiisotop Blei-206. Eine andere Zerfallsreihe beginnt beim Uranium-235 und endet beim Blei-207.
Das erste Element der Reihe, die Muttersubstanz, liefert der Reihe immer neue Atome. Wenn sich die Muttersubstanz nicht merklich verringert, bildet sich innerhalb der Reihe ein Gleichgewicht heraus, ein stationärer Zustand. Elemente mit großen Halbwertszeiten sind in größerer Menge vorhanden als Elemente mit kleineren Halbwertszeiten. Für die Menge N_i einer Atomsorte i innerhalb einer Reihe gilt: $N_i/T_{1/2i} = \text{const.}$ bzw. $N_i \cdot \lambda_i = C$.
Die von radioaktiven Elementen freigesetzten α-Teilchen

fangen sich Hüllelektronen ein, und es entstehen Heliumatome. Derart erzeugt 1 kg Radium im Jahr 167 cm^3 Helium.

5.3.3.3. Herstellung nicht in der Natur vorhandener Radionuklide. Mit langsamen und schnellen Reaktorneutronen lassen sich viele vielfältige nutzbringend einsetzbare Radionuklide erzeugen. So entsteht im thermischen Neutronenfluß aus normalem Phosphor das β^--aktive Isotop ^{32}P (^{31}P (n, γ) ^{32}P; $T_{1/2}$ des ^{32}P = 14,3 Tage).

Einen Elementwandel erreicht man mit n, p- und n, α-Reaktionen, z.B. aus Schwefel, Phosphor, ^{32}S (n, p) ^{32}P, welches sich dann chemisch gut abtrennen läßt. Da sich Isotope chemisch genau wie die natürlichen Elemente verhalten, kann man sie z.B. Pflanzen oder Tieren über Dünge- bzw. Nahrungsmittel zuführen. Verfolgt man nun mit Detektoren den Weg der radioaktiven Isotope durch die Pflanzen oder Körper, erhält man Informationen über die Verwertung und den Einbau der Elemente. Berühmt wurde in der Medizin der Iod-Test zur Überprüfung der Schilddrüsenfunktion. Beim Test von Wellen, Lagern bzw. Robotergelenken werden Oberflächen der Verschleißteile durch Bestrahlung aktiviert. Das Spülmittel wird auf seine Aktivität hin analysiert, die ein Maß für den Abrieb darstellt.

γ-Strahlen radioaktiver Präparate sind gut zur Durchstrahlung dicker Materialschichten geeignet, z.B. zur Güteprüfung von Schweißnähten an Rohren. Co-60 liefert γ-Strahlen mit einer Energie von $\approx$ 1,2 und 1,3 MeV. Dieses Cobalt-Isotop ist mit seiner Halbwertszeit von 5,24 Jahren eine brauchbar langlebige Strahlenquelle. Co-60 wird in der Medizin zur Zerstörung karzinomer Gewebe eingesetzt (Cobalt-Kanone).

Radioaktive Isotope werden in derartiger Menge und Vielfalt verwendet, daß Kernreaktoren allein zur Isotopenproduktion rentabel arbeiten können.

Radiopharmaka sind mit radioaktiven Isotopen versehene, also radioaktive Arzneimittel. Ihr Weg durch den Körper kann mit kernphysikalischer Meßtechnik verfolgt werden. Die Verteilung des Radiopharmakons in dem zu untersuchenden Körperorgan oder -gewebe ist von seiner Funktion abhängig. Krankhafte Gewebepartien oder Organe sind so auffindbar. Biologische Prozesse werden durch die geringe Strahlenbelastung nicht merklich beeinflußt. Die Art der Strahlenbelastung ist mit der beim Röntgen vergleichbar, aber in der überwiegenden Zahl der Fälle wesentlich geringer. Das derzeit günstigste Radionuklid für die in-vivo-Diagnostik ist ein Isotop des Technetiums (^{99m}Tc). Dieses Isomer des Tc-99 hat eine Halbwertszeit von 6 Stunden. Die Aktivität hält einerseits lange genug an, um den Weg des Arzneimittels durch den Körper bzw. seine Verteilung in einem bestimmten Gebiet analysieren zu können, andererseits klingt sie schnell genug ab, so daß die Strahlenbelastung nicht unnötig groß ist. Das Isomer geht unter Aussendung eines γ-Quantes der Energie W_γ = 0,14 MeV in den Grundzustand über. Diese Energie ist ausreichend, die Strahlung außerhalb des Körpers gut nachweisen zu können. Das Isotop im Grundzustand ist langlebig ($2 \cdot 10^5$ Jahre), es tritt keine merkliche sonstige Strahlenbelastung auf. Das Isomer kann in den Kliniken für Nuklearmedizin aus im Kernreaktor hergestelltem Mo-99 selbst gewonnen werden. Man schätzt, daß heutzutage 85% aller Radiopharmaka mit ^{99m}Tc markiert sind. Entsprechend der verschiedenen Anwendungsfälle wird es in die geeigneten chemischen Verbindungen eingebaut. Das nächstgünstigere Isotop I-123 hat eine Halbwertszeit von 13 Stunden. Mit der weiteren Entwicklung der Nachweistechnik werden kurzlebigere Isotope eingesetzt. Sehr günstig wären C-11 bzw. N-13 mit Halbwertszeiten von 20 bzw. 10 min. Diese Isotope haben den Vorteil, in die Medikamente wesentlich besser einbaubar zu sein, da Kohlenstoff und Stickstoff sowieso Bestandteile eingesetzter Pharmaka sind (Tab. 5.3.3–4).

Tab. 5.3.3–4 Häufig genutzte Radioindikatoren

Nuklid	Halbwertszeit	Strahlungsenergie in MeV und Strahlungsart
Na-24	15,0 h	1,39 β^- 2,76 γ
P-32	14,3 d	1,71 β^-
Mn-52	5,7 d	0,57 γ 1,46 γ
Fe-59	45,1 d	0,46 β^- 1,10 γ
Br-82	35,9 h	0,44 β^- 0,55 γ
Te-99m	6,0 h	0,14 γ
In-113m	1,7 h	0,39 γ
I-131	8,1 d	0,61 β^- 0,36 γ
Xe-133m	2,3 d	0,23 γ
La-140	40,2 h	1,36 β^- 1,60 γ
Au-198	2,7 d	0,96 β^- 0,41 γ

5.3.3.4. Altersbestimmung mit Radionukliden. Je nachdem, ob man Jahrtausende, Jahrhunderte oder Jahre mit der Altersbestimmung erfassen will, ist diese oder jene Methode anwendbar. Die Halbwertszeiten der zu verwendenden Radionuklide müssen dem Meßproblem angepaßt sein. Das Zerfallsgesetz (vgl. 5.3.3.2.) enthält die Zeit t, die vom Beginn des radioaktiven Zerfalls bis zum Meßzeitpunkt vergangen ist, und die Menge der radioaktiven Atome zu den beiden Zeitpunkten.

Uranium-Blei-Methode. Enthält ein Mineral Uranium, so wird der als Beginn des Zerfalles anzusetzende Zeitpunkt die Bildung der festen Erdkruste sein. Die Halbwertszeit des U-238 beträgt $4{,}51 \cdot 10^9$ Jahre. U-238 wandelt sich über eine Reihe von Zerfällen in das stabile Endprodukt Pb-206 um.

Mißt man die Menge des im Mineral entstandenen Pb-206 und die Restmenge Uranium-238, so ergibt sich das Alter des Minerals zu $t = (1/\lambda) \ln (N_{Pb}/N_u + 1)$.

Alle Messungen führen zu dem gleichen Ergebnis für den Zeitraum, der seit der Bildung der festen Erdkruste vergangen ist, von $\approx 4 \cdot 10^9$ Jahren.

C-14-Methode. Neben den stabilen Kohlenstoffisotopen C-12 (98,88%) und C-13 (1,12%) existiert auf der Erde eine geringe Gleichgewichtsmenge des radioaktiven Kohlenstoffisotopes C-14, welches mit einer Halbwertszeit von 5568 Jahren zerfällt. Dieser „Radiokohlenstoff" entsteht in der Hochatmosphäre durch über die Höhenstrahlung entstandene schnelle Neutronen aus Stickstoff ^{14}N (n, p) ^{14}C. Man schätzt die in der Lufthülle enthaltene Gleichgewichtsmenge an $^{14}CO_2$ auf $\approx$ 80 t. Durch das CO_2 der Luft assimilieren Pflanzen auch einen entsprechenden Anteil C-14. Jedes Lebewesen baut über die Nahrung C-14 ein, der Mensch nimmt im Laufe seines Lebens $\approx 10^{-18}$ g C-14 auf.

Mit Absterben der Pflanzen bzw. dem Lebensende der Tiere und Menschen ist die C-14-Aufnahme beendet und die Menge beginnt sich durch den radioaktiven Zerfall zu verringern – die Atomuhr ist eingeschaltet. Die C-14-Methode zur Altersbestimmung ist zu hoher Perfektion entwikkelt. Mit ihr läßt sich das Alter eines ≈ 800 Jahre alten Holzbootes genauso gut bestimmen, wie das ägyptischer Königsmumien von einigen tausend Jahren.

Tritium-Methode. Analog dem C-14 wird das schwere Wasserstoffisotop ^3H (auch mit T bezeichnet) – Tritium –, gebildet. Es ist in geringer Menge im Wasser als T_2O enthalten. Die Halbwertszeit von 12,3 Jahren macht es für Altersbestimmungen bis zu einigen zehn Jahren brauchbar.

5.3.4. Kernspaltung

5.3.4.1. Die Entdeckung der Kernspaltung. Wie auch andere Physiker der damaligen Zeit, versuchte Joliot-Curie durch den Beschuß schwerer Kerne mit Neutronen noch schwerere Kerne herzustellen. Neutronen gelangen leicht in den Bereich der Kernkräfte, da sie als ungeladene Teilchen nicht gegen die Coulomb-Abstoßung der Protonen anlaufen müssen. Als Neutronenquelle wurde damals ein Gemisch aus Radium und Beryllium verwendet. Radium strahlt α-Teilchen ab, die mit Beryllium zu Kohlenstoff unter Aussendung von Neutronen reagieren. Diese Neutronen haben Energien von mehr als 10 MeV. Meist ist es günstiger, sie erst abzubremsen, zu verlangsamen, da sie sich dann länger in Kernnähe, im Bereich der Kernkräfte aufhalten, was die Reaktionswahrscheinlichkeit erhöht. Zum Abbremsen eignet sich Paraffin. Joliot-Curie bestrahlte Uranium mit Neutronen, um Elemente schwerer als Uranium (Transurane) herzustellen. Die chemische Analyse ergab aber ein Element (Lanthan) mit den Eigenschaften der seltenen Erden. Hahn, Meithner und Strassman analysierten außerdem einwandfrei das Element Barium. Einzige Erklärung war die damals nicht vermutete und nicht erklärbare Spaltung des Uraniums nach der Reaktion $^{235}_{92}U + n \rightarrow {}^{142}_{56}Ba + {}^{91}_{36}Kr + 3n$, denn auch das Krypton konnte chemisch nachgewiesen werden. Danach dauerte es nur noch acht Jahre, bis die ersten beiden über Hiroshima und Nagasaki abgeworfenen Atombomben über 100000 Menschen das Leben kosteten, aber auch der erste Kernreaktor 1954 die friedliche Nutzung demonstrierte.

5.3.4.2. Der Spaltprozeß. Nicht jeder, auch nicht jeder schwere Kern ist in der einleitend beschriebenen Art spaltbar. Wenn es gelingt, dem schweren Kern ein Neutron anzulagern, wird die Bindungsenergie dieses Neutrons frei. Verbleibt die Energie von mehr als 5 MeV im Kern, ist er damit in der Lage, Formschwingungen auszuführen. Der Kern nimmt eine gestreckte Form an, schwingt zurück, streckt sich wieder usw. Die Schwingung kann sich so weit aufschaukeln, daß regelrecht zwei Hälften mit beginnender Einschnürung dazwischen existieren. In der Situation wirken starke auseinander treibende Kräfte, die zur Spaltung führen. Die beiden durch ihre Protonen positiv geladenen Hälften stoßen sich gegenseitig ab – die Spaltung ist perfekt. Abb. 5.3.4–1 erklärt diesen Prozeß. Den Vorstellungen liegt das Modell des Flüssigkeitstropfens zugrunde (vgl. 5.2.3.4.). Die Formschwingungen führen zu größer werdendem Abstand R der Schwerpunkte der Teilbereiche. Die Teilbereiche schwingen so lange zurück, bis der kritische Abstand R_{Sp} erreicht ist. Die Schwingungsenergie ist dann gleich der *Aktivierungsenergie*, der Energie, die notwendig ist, die Spaltbarriere ΔW zu überwinden. Die Bindungsenergie des angelagerten Neutrons kann bis ≈ 7 MeV betragen, nur Kerne mit nicht wesentlich größerer Aktivierungsenergie sind als leicht spaltbar zu bezeichnen. Da die Aktivierungsenergie durch die Abstoßungskräfte der Protonen verringert wird, ist in der Regel ein Kern um so leichter spaltbar, je schwerer er ist. Die Aktivierungsenergie ist kleiner als 10 MeV für Kerne mit $A \gtrsim 230$. Ab $A \approx 250$ ist keine Aktivierungsenergie mehr notwendig, die Spaltbarriere ist Null. Derartige Kerne können nur kurzzeitig existieren und spalten spontan. Kerne mit einer nicht zu hohen Spaltbarriere, wie das U-235, spalten auch spontan, d. h. ohne die

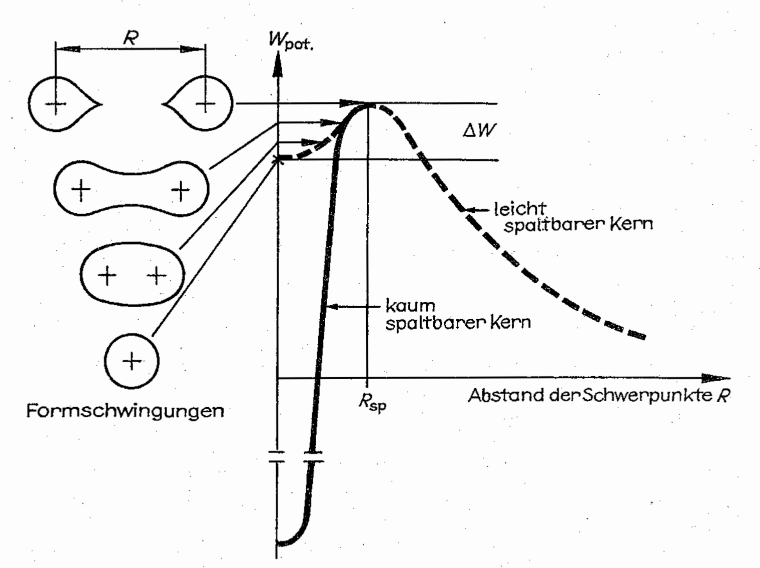


Abb. 5.3.4–1 Potentialverlauf leicht- und schwerspaltbarer Kerne sowie die Aufnahme von Schwingungsenergie bis zur Spaltbarriere ΔW

Bindungsenergie eines Neutrons zu benötigen. Die Spaltung erfolgt über den quantenmechanischen Tunneleffekt, analog dem α-Zerfall. Die spontane Spaltung des U-235 hat eine Halbwertszeit von 10^{17} Jahren und ist ein damit sehr seltenes Ereignis, wesentlich unwahrscheinlicher als der α-Zerfall.

Spaltmaterialien. Materialien, bei denen die Bindungsenergie eines angelagerten Neutrons ausreicht, den Kern zu spalten, sind U-235, U-233 und Pu-239. Bringt das Neutron neben der Bindungsenergie noch eine kinetische Energie von 1,5 bzw. 2 MeV in die Reaktion mit ein, so sind auch U-238 bzw. Th-232 spaltbar. Die Aktivierungsenergie zur Spaltung liegt um diese 1,5 bis 2 MeV über der Bindungsenergie eines angelagerten Neutrons. Somit kann man zwischen Spaltungen mit schnellen, energiereichen und langsamen Neutronen unterscheiden. Genügen langsame Neutronen zur Spaltung, versucht man auch möglichst langsame zu erzeugen, da damit die Spaltung besonders wahrscheinlich wird. Der technisch einfachere Prozeß ist der der Spaltung mit langsamen Neutronen. Von den genannten Spaltmaterialien sind das Uraniumisotop 238, das Uraniumisotop 235 und das Thoriumisotop 232 natürlich vorkommend. Natururanium besteht im wesentlichen aus U-238 und nur zu 0,7 % aus dem so effektiv zur Spaltung zu verwendenden U-235. Plutonium-239 und Uranium-233 sind über Kernreaktionen im Reaktor zu erzeugen, zu erbrüten. Pu-239 entsteht im Reaktor aus U-238:

$$\text{U-238} + \text{n} \rightarrow \text{U-239} + \gamma \xrightarrow{\beta^-} \text{Np-239} \xrightarrow{\beta^-} \text{Pu-239}.$$

Ähnlich wird im Reaktor U-233 erbrütet, Ausgangsmaterial ist das Th-232:

$$\text{Th-232} + \text{n} \rightarrow \text{Th-233} \xrightarrow{\beta^-} \text{Pa-233} \xrightarrow{\beta^-} \text{U-233}.$$

Spaltenergie. Die bei der Spaltung schwerer Kerne freiwerdende Energie ist schon mit ≈ 200 MeV je Spaltung abgeschätzt (vgl. 5.2.3.3.). Eine andere einfache Betrachtung führt zum gleichen Ergebnis. Im Moment der Entstehung der Spaltbruchstücke wirken die auseinander treibenden Coulomb-Kräfte und beschleunigen die Teile bis zu einer kinetischen Energie von

$$W^{\text{kin}}_{\text{Spalt}} = (Z_1 \cdot Z_2 e^2)/[4\pi\varepsilon_0 (R_1 + R_2)] = 166 \text{ MeV}.$$

Z_1 und Z_2 sind die Ladungen der Spaltbruchstücke, R_1 und R_2 ihre Radien und $R_1 + R_2$ der Abstand beider Schwerpunkte im Moment der Spaltung. Zu dieser Energie gehen in die Bilanz noch ein: 5 MeV in Form von kinetischer Energie auch emittierter Neutronen, 20 MeV aus nachfolgenden Zerfällen der Bruchstücke, 10 MeV gehen mit Neutrinos verloren; aufgrund der geringen Wechselwirkungsfähigkeit geben sie ihre Energie nicht so bald an die Umgebung ab. Die überwiegende Menge der freiwerdenden Energie ist kinetische Energie. Die energiereichen Teilchen werden im Reaktor abgebremst, die entstehende Wärme wird über die üblichen Turbinen (der Öl- bzw. Kohlekraftwerke) in Elektroenergie umgewandelt.

5.3.4.3. Die Kettenreaktion. Die Kernspaltung wird durch Neutronen ausgelöst. Man benötigt $\approx 10^{17}$ Neutronen für 1 KWh elektrischen Strom. Kleine Stromquellen für die Versorgung von Bojen im Ozean oder für Satelliten kann man mit Neutronenquellen und Uranium im Gemisch betreiben, Kernreaktoren arbeiten anders. Bei der Spaltung des Kernes entstehen neben den zwei Bruchstücken noch zwei bis drei einzelne Neutronen, da die Bruchstücke für ihre Massenzahl überreich an Neutronen sind. Auch nach der Emission dieser prompten Spaltneutronen haben die Teilstücke noch zu viele Neutronen, so daß in nachfolgenden Zerfällen die Teile sich auch noch durch Emission von Neutronen – den sogenannten verzögerten Neutronen – stabilisieren. Verzögerte Neutronen sind für die Reaktorsteuerung von besonderer Bedeutung.

Mit beiden Sorten der bei der Kernspaltung freiwerdenden Neutronen können wieder Spaltungen eingeleitet werden. Wenn bei einer Spaltung mehr Neutronen entstehen, als für die nächste benötigt werden, steigt die Zahl der Spaltungen lawinenartig an. Wenn eine erste Spaltung zwei neue einleitet und davon wieder jede zwei neue, dann ist sehr rasch alles Material gespalten. Die lawinenartige Vermehrung der Spaltrate nennt man *Kettenreaktion.*

Eine ungesteuerte, frei ablaufende Kettenreaktion spaltet in einigen Millisekunden einige Kilogramm Uranium (Kernspaltungsbombe).

Kritische Massen. Voraussetzung für die Kettenreaktion ist, daß die Spaltneutronen wiederum spalten. Dazu müssen einige Bedingungen erfüllt sein. Die Neutronen dürfen weder durch andere Reaktionen verloren gehen noch ins Freie entweichen können. Eine Möglichkeit ist, so viel leicht spaltbares Material zu nehmen (U-235), daß die Neutronen nur auf Spaltmaterial treffen können. Selbst dann ist die Spaltung noch ein so seltenes Ereignis, daß im Mittel erst nach einer Weglänge von 17 cm in U-235 bzw. nach 10 cm in Pu-239 wieder eine Spaltung stattfindet. Hat man so viel Spaltmaterial, daß die Weglängen zwischen zwei Spaltungen im Material liegen, setzt die Kettenreaktion ein. Der derart mindestens benötigte Durchmesser für eine kugelförmige Materialmenge legt die kritische Masse fest. Die kritische Masse ist die Menge Spaltstoff, die zum Ablauf einer Kettenreaktion mindestens benötigt wird.

Die kritische Masse beträgt ≈ 50 kg für reines U-235. Die kritische Masse kann z. B. durch eine Hülle neutronenreflektierenden Materials verringert werden. Gegenwärtig scheint die Militärtechnik in der Lage zu sein, kritische Massen in der Größenordnung 10^{-2} kg realisieren zu können. Wird für den Reflektor Beryllium eingesetzt, so arbeitet dieser noch über die Reaktion ^{9}Be (n, 2n) ^{8}Be als Neutronenvervielfacher. Dadurch verringert sich beispielsweise die kritische Masse des U-235 auf 15 kg, die des Plutoniums auf 4,4 kg. Es existieren auch einige seltene bzw. schwer herzustellende Elemente mit sehr großem Spaltquerschnitt und einer größeren Anzahl von Spaltneutronen. Bestimmte Isotope des Curiums und des Californiums haben kritische Massen von ≈ 25 g.

Könnte man Massen $\lesssim 10^{-6}$ kg zünden, ließe sich damit direkt mechanische Energie in der Art der Verbrennungskraftmaschinen erzeugen.

5.3.4.4. Die gesteuerte Kernspaltung. Die Kettenreaktion kann so gesteuert werden, daß die Zahl der Spaltungen in der Sekunde gleich bleibt. Zur Beschreibung dieses Vorganges führt man die Größen „Vermehrungsfaktor" K und „Lebensdauer einer Neutronengeneration" τ ein. τ entspricht der Zeit zwischen der Spaltung, in der es entsteht und der, in der es die neuen Neutronen erzeugt. Mit K und τ ergibt sich die Zahl der Spaltungen zu $N = N_0 \exp[(k-1)t/\tau]$. N_0 ist die Zahl der Spaltungen zu Beginn der Betrachtungen, etwa beim Anlaufen eines Reaktors die erste Spaltung, oder im Betrieb die Zahl der notwendigen Spaltungen. Für $K > 1$ steigt die Zahl der Spaltungen exponentiell an, für $K < 1$ erlischt der Reaktor. Der normale Reaktorbetrieb erfolgt mit $K \approx 1$. τ beträgt z. B. für 2-MeV-Neutronen 10^{-8} s. Man könnte annehmen, daß es sehr schwer ist, derart schnell ablaufende Reaktionen zu steuern, kon-

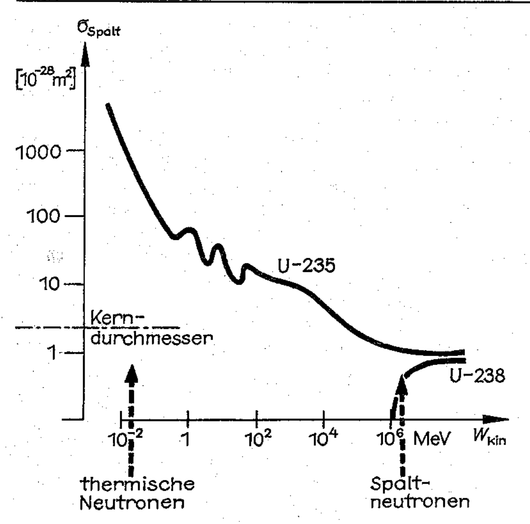


Abb. 5.3.4–2 Skizze der Wirkungsquerschnitte für die Spaltung von U-235 und U-238 mit Neutronen in Abhängigkeit von ihrer Energie

trolliert ablaufen zu lassen. Ein glücklicher Umstand vereinfacht die Reaktorregelung wesentlich. Die Neutronenbilanz enthält auch die in den Nachfolgezerfällen entstehenden verzögerten Neutronen (10 bis 20 s nach der Spaltung). Für einen Vermehrungsfaktor von $K \leqq 1{,}003$ ist die Zeit, in der der Reaktor geregelt werden muß (Verdopplung bzw. Halbierung der Zahl der Spaltungen) größer als 10 s, weil die Regelung mit den unbedingt noch benötigten verzögerten Neutronen arbeitet. Erst ab Vermehrungsfaktoren $K > 1{,}0075$ genügen die prompten Neutronen zur Aufrechterhaltung der Spaltungsmenge, muß der Reaktor innerhalb 10^{-7} s geregelt werden können. Normale Reaktoren sind so ausgelegt, daß sie immer die 0,75% verzögerten Neutronen benötigen und nicht prompt kritisch werden können.

Wirkungsquerschnitte für die Kernspaltung. Die bei der Kernspaltung freiwerdenden Neutronen haben Energien der Größenordnung 1 MeV. Wirkungsquerschnitte für die Spaltung in Abhängigkeit von der Neutronenenergie zeigt Abb. 5.3.4–2. Die Abbildung verdeutlicht, daß mit MeV-Neutronen die Kernspaltung ein seltenes Ereignis ist, daß der Wirkungsquerschnitt nur ≈ 1/500 des Wirkungsquerschnittes für thermische Neutronen ist. Thermische Neutronen haben eine Energie, die umgerechnet der Zimmertemperatur entspricht (≈ 25 meV). Die Spaltquerschnitte für schon mit thermischen Neutronen spaltbare Stoffe (U-235, U-233, Pu-239) sind alle etwa gleich. Wird zur Spaltung mehr als die freiwerdende Bindungsenergie des Neutrons benötigt (U-238, Th-232), so setzt die Spaltung erst bei ausreichender kinetischer Energie des Neutrons ein. Um effektiv spalten zu können, benötigt man thermische Neutronen, muß man die schnellen Neutronen abbremsen, moderieren (vgl. 5.3.1.3.).

5.3.5. Kernspaltungsreaktoren

Je nach Art der Arbeitsweise unterscheidet man zwischen den derzeit überwiegend im Betrieb befindlichen *thermischen Reaktoren* und den noch seltenen, aber für die Zukunft wesentlichen *schnellen Reaktoren.* Letztere werden wegen ihrer Eigenschaft, auch Brennmaterial für thermische Reaktoren zu erzeugen, zu erbrüten, auch *Brutreaktoren* genannt.

Die Bezeichnung thermisch bzw. schnell bezieht sich auf die Eigenschaft der die Spaltung überwiegend auslösenden Neutronen, schnell oder langsam (thermalisiert) zu sein. Schnelle Neutronen haben Energien von ≳ 1 MeV und damit Geschwindigkeiten um $10^4\,\text{km/s}^{-1}$. Thermalisierte Neutronen weisen Energien um 0,025 eV (entsprechend 300 K) auf, die dazugehörigen Geschwindigkeiten betragen $\approx 2\,\text{km/s}^{-1}$.

Thermische Reaktoren lassen sich wesentlich einfacher bauen und betreiben als schnelle Reaktoren. Thermische Reaktoren sind technisch ausgereift und in vielen Staaten der Erde schon jahrzehntelang in Betrieb, wohingegen sich die schnellen Brüter noch in der Erprobung befinden und bei wesentlichen technologischen Problemen wird noch die optimale Lösung gesucht.

5.3.5.1. Thermische Reaktoren. Prinzipielle Arbeitsweise.

Abb. 5.3.5–1 veranschaulicht die Wirkungsweise eines Kernkraftwerkes mit thermischem Reaktor. Es gliedert sich in die Baugruppen Reaktor, primärer Kühlkreislauf mit Wärmeaustauscher und sekundärer Kühlkreislauf mit Turbine und Generator. Ab Sekundärkreislauf wird die konventionelle Technik der Kohlekraftwerke verwendet. Reaktor und der auch hochradioaktive primäre Kreislauf zur Wärmeabführung sind für eventuelle Havarien gegen das Entweichen gasförmiger, flüssiger und fester radioaktiver Substanzen in einer drucksicheren Schale aus Stahl oder Beton untergebracht. Der eigentliche Reaktor besteht in der hier dargestellten Ausführungsform aus den leicht zu wechselnden Brennstoffstäben, die in eine Moderatorsubstanz (Bremsmaterial) eingelassen sind und dem stark Neutronen absorbierenden Regelstab. Die Wärmeabführung kann verschiedenartig ausgeführt werden, z. B. mit Kühlrohren im Moderatorblock.

Neutronenbilanz. Wesentlich für den Betrieb eines Reaktors ist der Vermehrungsfaktor K (vgl. 5.3.4.4.). Er gibt das Verhältnis der Zahl der Spaltungen zweier aufeinanderfolgender Neutronengenerationen an und ist gleich dem Verhältnis der Neutronenzahlen zweier aufeinanderfolgender Neutronengenerationen. Der stationäre Betrieb eines Reaktors erfordert $K = 1$, damit der Reaktor regelbar ist, muß ein $K > 1$ einstellbar sein. Ob ein Reaktor überhaupt kritisch werden, d. h. in den stationären Betrieb kommen kann, kann mit der *Vierfaktorenformel* für den Multiplikationsfaktor abgeschätzt werden. Prinzipiell ist die Zuwachsrate (K-1) für die Neutronen positiv, da bei der durch ein Neutron erfolgenden Kernspaltung zwei bis drei Neutronen entstehen.

Mit der Vierfaktorenformel berechnet sich für eine unendlich ausgedehnte Reaktionszone der Multiplikationsfaktor K zu $K_\infty = \eta \cdot \varepsilon \cdot p \cdot f$.

Für praktische Reaktorgrößen muß noch die Neutronenleckrate (1-P) berücksichtigt werden, $K = K_\infty \cdot P$. P ist der Quotient aus der Zahl der in der Reaktionszone verbleibenden und der Zahl der dort erzeugten Neutronen.

η ist eine effektive Spaltneutronenzahl. Die eigentliche Zahl der erzeugten Spaltneutronen je in der Spaltsubstanz absorbiertem Neutron ist größer, die effektive Neutronenzahl entsteht durch Berücksichtigung des Einfanges ohne Spaltung auch im U-238, wenn es sich um angereichertes U-235 handelt. η beträgt für Natururanium (0,71% U-235) 1,35, bei 10% U-235 2,00 und erreicht für 100% U-235 den

Wert 2,08. ε, ein Faktor etwas größer als 1, z. B. 1,02 gibt den Neutronengewinn durch Spaltung ohne Moderation der Neutronen (mit schnellen Neutronen) an. p ist kleiner als 1, es berücksichtigt den Neutronenverlust bei der Moderation, 0,9 ist ein typischer Wert. f ist der Faktor der Nutzung thermischer Neutronen. Er wird gebildet aus dem Verhältnis der im Spaltstoff absorbierten thermischen Neutronen zu denen im Material allgemein absorbierten thermischen Neutronen ($f \approx 0{,}9$). Ein Reaktor muß mit einiger Reserve für den Multiplikationsfaktor K berechnet werden, da sich K während des Betriebes verringert, auch entstehen als Spaltprodukte im Brennstoff stark Neutronen absorbierende Substanzen.

Kritische Massen. Genaue Betrachtungen aller im Reaktorkern wirkenden Faktoren führen zu Mindestmengen an Kernbrennstoff, ab der die Reaktoren kritisch werden (arbeiten) können. Diese vom Reaktortyp abhängige Größe wird *kritische Masse* genannt. Ein thermischer Reaktor mit reinem U-235 kann ab 1 kg Brennstoff kritisch werden, bei auf 5% angereichertem U-235 und Graphit-Moderator liegt die kritische Masse bei 250 kg. Ein mit Uranium natürlicher Zusammensetzung arbeitender und mit schwerem Wasser moderierender Reaktor benötigt über 2000 kg Uranium, ist der Moderator Graphit, sind mehr als 40000 kg Mindestmenge Uranium notwendig.

Moderatorsubstanzen. Je leichter die Atome einer Substanz sind, je schneller bremsen sie Neutronen ab (vgl. 5.3.1.3.). Notwendige Eigenschaft einer Moderatorsubstanz muß aber auch eine geringe Neutronenabsorption sein. Das berücksichtigend ist die, wenn auch sehr teure günstigste Substanz „schweres Wasser" (D_2O). Noch sehr günstig zur Moderation ist Graphit zu verwenden (synthetisch hergestellt, von höchster Reinheit). Ihm folgen in der Reihenfolge Beryllium und Wasser (H_2O).

Prähistorischer Kernreaktor. Uranium hat auf der Erde ein konstantes Isotopenverhältnis von 0,7% U-235 zu 99,3% U-238. Vor $\approx$ 2 Milliarden Jahren muß das Isotopenverhältnis noch etwa 3% zu 97% betragen haben. In der Zeit kann in einem Uraniumerzlager durch eindringendes Wasser moderiert, eine Kettenreaktion in Gang gesetzt worden sein. Durch die Spaltungswärme konnte dann wieder soviel Moderatorwasser verdampfen, daß die Reaktion regelrecht gesteuert ablief, bis der natürliche Isotopenanteil soweit abgesunken war, daß die Kettenreaktion zum Erliegen kam. Der hier geschilderte Vorgang ist die Erklärung für einen in einer Uraniumlagerstätte in Afrika (Gabun) festgestellten anormal geringen Isotopenanteil von 0,44% U-235.

Energieumsetzung in Kernreaktoren. Eine große Anzahl von Kernreaktoren, zumeist mit kleinerer Leistung, werden nicht zur Energieerzeugung direkt gebaut und eingesetzt. Noch werden Reaktoren für Forschungs- und Ausbildungszwecke benötigt, insbesondere ist die dort auch angesiedelte Isotopenproduktion von hohem Wert. Für Forschung und Isotopenproduktion benötigt man hohe Neutronenflüsse – Kanäle in den Reaktor zu Orten hoher Neutronendichte. Auch aus diesen Reaktoren muß die Wärme abgeführt werden, wobei es sich auch um Leistungen im MW-Bereich handelt.

Kernkraftwerke werden überwiegend zur Elektronenenergieerzeugung eingesetzt. Sie arbeiten so sicher, daß sie sowohl in dicht besiedelten Gebieten wie auch an den entlegendsten Stellen der Erde errichtet werden können. Ihr großer Vorteil ist einmal, daß sie statt z.B. 10^6 kg Steinkohle je Tag nur 1 kg U-235 benötigen und zum anderen die direkte Umweltverschmutzung nur in Form von Kühlwassererwärmung zu berücksichtigen ist. Es ist üblich, den Wirkungsgrad von Kernkraftwerken mit 25% anzugeben. Das ist aber nur der Wirkungsgrad der Werke selbst. Beschreiben kann man einen Kernreaktor durch die aus den Kernreaktionen resultierende thermische und abgegebenen elektrischen Leistung, die $\approx 1/4$ der thermischen Leistung beträgt. Wesentlich ist noch zu wissen, bis zu wieviel Prozent das Uranium im Reaktor ausgenutzt werden kann, ehe der Vermehrungsfaktor $K < 1$ wird und die Kettenreaktion ab-

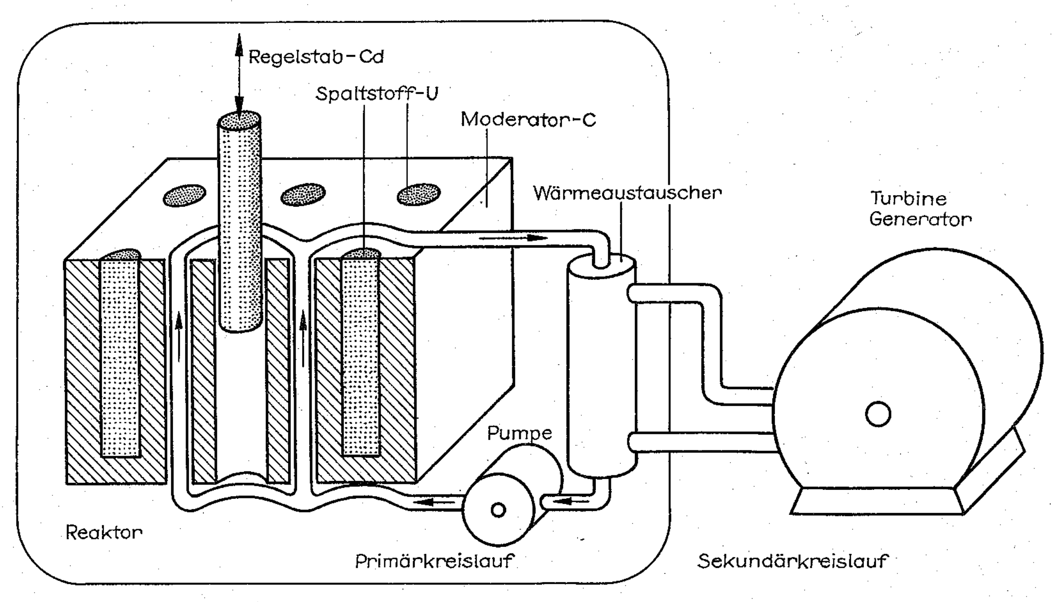


Abb. 5.3.5-1 Prinzipbild eines Kernkraftwerkes mit thermischem Reaktor

bricht. Der Wert liegt in der Regel bei 50%. Wesentlich ist auch zu berücksichtigen, unter welchem Energieeinsatz die Anreicherung des U-235-Isotopes auf z. B. 10% erfolgt usw. Kalkulationen in kapitalistischen Ländern weisen aus, daß unter deren Bedingungen Kernkraftwerke derzeit schon rentabel arbeiten. Aus der Sicht der Physik ist bei der Umstellung auf andere Energieträger, als die sich erschöpfenden fossilen Brennstoffe, die Nutzung der Kernenergie notwendig und auf dem derzeitigen technischen Stand bedenkenlos möglich.

Typen von Kernreaktoren. Wenn Spalt- und Moderationsmaterial homogen im Reaktor vorliegen, handelt es sich um einen homogenen Reaktor, in den meisten Fällen sind die Brennstäbe in eine Moderatorsubstanz eingelassen, eine heterogene Anordnung, die den heterogenen Reaktor charakterisiert. In der Abb. 5.3.5–2 ist ein homogener Reaktor in Form zweier konzentrischer Kugeln dargestellt. Die äußere Kugel wird von einer Kühlflüssigkeit umspült, sie ist noch von einem stabilen Schutzmantel umgeben. Die innere Kugel bildet die aktive Zone. Ab 30 cm Durchmesser wird sie mit einer wässrigen Uraniumsalzlösung kritisch. Die Brennstoff-Moderator-Flüssigkeit kann kontinuierlich aufgefrischt werden, bei Überhitzung entweicht die Lösung durch einen Überlauf und die Kettenreaktion erlischt. Dieser homogene Reaktor wird als „Wasserkocher" bezeichnet.

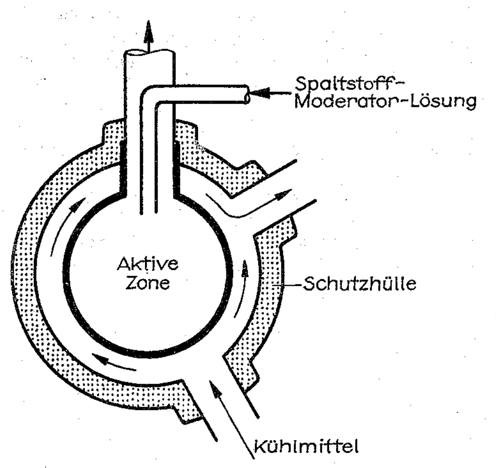


Abb. 5.3.5–2 Homogener Kernreaktor

Bei der radiolytischen Zersetzung des Wassers entstehen Gasblasen, die ein Geräusch verursachen, als wenn Wasser kocht. Insbesondere die heterogenen Reaktoren werden noch durch eine ausführliche Schreibweise untergliedert, in der Spaltmaterial, Moderatorsubstanz und Kühlmittelart angegeben sind, z. B. Natururanium-Graphit-Gas-Reaktor, angereicherter Uranium-Wasser-Wasser-Reaktor. Bei wassergekühlten Reaktoren unterscheidet man noch zwischen Druck- und Siedewasserreaktoren. Kommt die Kühlflüssigkeit in der aktiven Zone zum Sieden, handelt es sich um Siedewasserreaktoren, wird das Sieden durch einen Druckbehälter verhindert, spricht man von einem Druckwasserreaktor. Druckwasserreaktoren sind wegen ihrer kompakten Bauweise (Arbeitstemperatur z. B. 320 °C, Arbeitsdruck z. B. 15,2 MPa) als Schiffsantriebe geeignet.

5.3.5.2. Brutreaktoren. Kernreaktoren verbrauchen nicht nur Spaltstoff, sondern sind auch über zwei Reaktionen in der Lage, neuen Spaltstoff zu erzeugen. Erzeugt der Reaktor mehr Spaltstoff als er verbraucht, wird er zum Brutreaktor. Unter Spaltstoff in diesem Sinne versteht man das mit thermischen Neutronen spaltbare Material U-233 und Pu-239. Das Plutonium entsteht über den Einfang schneller Neutronen durch das häufige Uraniumisotop U-238 nach zwei nachfolgenden β^--Zerfällen. U-233 wird aus in ausreichender Menge auf der Erde vorhandenem Thorium-232 erzeugt, auch über Neutroneneinfang und zwei nachfolgenden β^--Zerfällen. Ein wenig brüten alle thermischen Reaktoren, da sie in genügender Menge U-238 – in dem der Spaltstoff U-235 nur angereichert ist – in der aktiven Zone aufweisen. Jedoch sind die Verluste durch Moderation und Kühlung zu groß, um mehr Spaltstoff zu erzeugen, als verbraucht wird. Der Bau von Brutreaktoren gelingt nur ohne Moderation, also über Spaltreaktionen mit schnellen Neutronen.

Brutrate, Brutgewinn. Der Konversions- oder Brutfaktor K gibt die Zahl der erzeugten spaltbaren Atome je Spaltung an, $K = \eta - 1 -$ Verluste. K ergibt sich aus der effektiven Spaltneutronenzahl η durch Subtraktion eines Neutrons zur Aufrechterhaltung der Spaltung im Reaktor und Abzug der Neutronenverluste durch z. B. Moderation. Zieht man noch einmal ein Neutron zur Selbstversorgung des Reaktors mit Spaltmaterial ab, verbleibt der Brutgewinn $= K - 1$.

In thermischen Reaktoren ist kein Brutgewinn erzielbar, da die effektive Spaltneutronenzahl η nur $\approx 2{,}1$ beträgt und der Neutronenverlust je Spaltung mit 0,3 anzusetzen ist. Bei Spaltung mit schnellen Neutronen ist die effektive Spaltneutronenzahl größer und für Plutonium mit $\eta = 2{,}7$ ist theoretisch ein Brutgewinn von 0,7 erreichbar. Reaktoren mit schnellen Neutronen sind zwar technisch erheblich komplizierter als thermische Reaktoren, aber zur Sicherung der Energieversorgung günstig, da das auf der Erde ausreichend vorhandene U-238 und Th-232 genutzt werden kann. Reaktoren, die mit schnellen Neutronen spalten und einen Brutgewinn erreichen, werden *schnelle Brüter* genannt.

Brutreaktoren bestehen aus einem Spaltkern mit zumeist einem Gemisch aus U-235 und Pu-239, der von einem Brutmantel aus U-238 und/bzw. Th-232 umgeben ist. Problematisch ist die Kühlung und Regelung des Reaktors. Kühlmittel, Brennelementummantelungen u. a. Konstruktionselemente in und am Kern des Reaktors dürfen die Spaltneutronen weder moderieren noch stark absorbieren. Als Kühlmittel wird gegenwärtig flüssiges Natrium verwendet. Zur Reaktorregelung existieren keine ausgesprochen guten Absorber, wie das Cadmium für langsame Neutronen, eine Abstandsregelung wird notwendig. Zudem muß die Regelung sehr schnell erfolgen, da der Reaktor innerhalb 10^{-6} s seine Spaltrate vervielfachen kann.

5.3.5.3. Atommüll. Ein Reaktor erzeugt so viele Spaltprodukte, wie er als Spaltmaterial verbraucht. Ein Reaktor von 100 MW thermischer Energie verbraucht im Jahr ≈ 40 kg U-235. Spaltprodukte sind die zwei aus einem spaltbaren Kern entstehenden mittelschweren Nuklide. Sie verteilen sich auf die Massenzahlen $70 \lesssim A \lesssim 165$. Diese Nuklide sind in der Regel, da sie mit hohem Neutronenüberschuß entstehen, instabil. Sie zerfallen zumeist unter Aussendung von Elektronen (β^--Zerfälle) über mehrere Stationen, bis ein stabiles Element erreicht ist. Beim Zerfall tritt harte γ-Strahlung auf, deren Halbwertszeiten den gesamten möglichen Bereich bis zu Jahrhunderten umfassen. Da der Abbrand meist nur bis zur Nutzung von 50% des eingesetzten spaltbaren Materials möglich ist, muß das restliche Spalt-

material zurückgewonnen werden, und da immer aus dem U-238 auch das Spaltmaterial Plutonium gebildet wird, ist auch dieses zur Nutzung abzutrennen. Weiterhin können viele der entstandenen Radionuklide (z. B. in der Medizin und Technik) eingesetzt werden und sind derzeit auf anderen Wegen auch nur sehr aufwendig zu erzeugen. Das frische Abprodukt kann in automatischen Anlagen sofort aufbereitet werden, oder kann, damit die kurzlebige Aktivität erst einmal abklingt, ≈ 1/2 Jahr zwischengelagert werden, um dann erst zur Aufbereitung zu kommen. Derzeit ist eine totale Aufbereitung nicht möglich und sinnvoll, so daß schon mit einer Endlagerung begonnen wird. Auch die Endlagerung ist – wenn sie mit aller Sicherheit erfolgt – ein kompliziertes technologisches Problem. Die Kapselung des radioaktiven Materials muß für alle Nuklide dicht sein, selbstverständlich der Wärmeentwicklung und den Wirkungen aller zu erwartenden chemischen Reaktionen standhalten sowie keine Aktivität an die Umgebung abgeben. Wird am Ende seiner Lebensdauer ein Reaktor stillgelegt, so sind auch der aktive Kern und alle aktivierten Materialien zum *Atommüll* hinzuzurechnen. Die weitere Entwicklung kann es durchaus nützlich werden lassen, die derzeit endgelagerten Abprodukte der Kernkraftwerke so zu behandeln, wie die Halden des Erzbergbaues vergangener Zeiten, die mittlerweile nicht nur einmal zur Nutzung aufbereitet wurden.

Kernenergetik. Gegen die Nutzung der Kernenergie zur Deckung des Energiebedarfs der Menschheit ist aus physikalischer Sicht kein alternatives Konzept notwendig. Der Weltenergiebedarf hat noch auf lange Zeit steigende Tendenz, insbesondere wenn man davon ausgeht, daß die derzeitigen Entwicklungsländer in den kommenden Jahren einen Energieverbrauch realisieren, wie ihn die hochindustrialisierten Länder bisher benötigten. Selbst wenn sich die prognostizierte schnelle Steigerung des Energiebedarfes mit Verdopplungszeiten von 20 Jahren nicht ergibt, so gehen dennoch die fossilen Rohstoffe zur Energieerzeugung in absehbarer Zeit zur Neige bzw. erhöhen sich schnell die Kosten für ihre Bereitstellung.

Kernkraftwerke mit thermischen Reaktoren haben sich über Jahrzehnte bewährt und bilden eine gute Ausgangsbasis zur Entwicklung leistungsfähiger Brutreaktoren. Der alleinige Einsatz thermischer Reaktoren würde die Vorräte an dem erforderlichen Uranium-Isotop U-235 etwa gleichschnell wie die von Kohle, Erdöl und Erdgas erschöpfen. Die seit 30 Jahren in der Entwicklung befindlichen schnellen Brutreaktoren arbeiten schon in kostengünstigen großen Dimensionen. An ihnen werden die letzten offenen Probleme für die Herstellung und den Betrieb gelöst. Mit ihnen kann aus dem Natururanium um zwei Größenordnungen mehr Energie gewonnen werden, als mit thermischen Reaktoren. Die bisher erkundeten Weltvorräte sind damit auf lange Sicht ausreichend.

Natriumgekühlte schnelle Brutreaktoren haben gegenüber thermischen Reaktoren wesentlich höhere Investitionskosten. Zum Teil kompensiert sich dies über die geringeren Brennstoffkosten, auch werden sie durch den Übergang zu noch größeren Leistungseinheiten wirtschaftlicher. Reserven für den ökonomischen Betrieb von Kernkraftwerksanlagen sieht man noch in einer Erhöhung des derzeit maximalen Abbrandes der Brennelemente von 10 auf 15%. Auch die Wiederaufbereitung verbrauchter Brennelemente kann kostengünstig sein.

Kernkraftwerke liefern die Wärmeenergie in dem von Verbrennungskraftwerken her beherrschten Temperaturbereich zwischen 300 und 550°C. Der Vorteil liegt im Einsatz erprobter Heißdampfverwerter, insbesondere von Turbinen. Die in der Entwicklung befindlichen Hochtemperaturreaktoren erhöhen nicht nur den Wirkungsgrad für die Elektroenergieerzeugung. Bei Kühlmittelaustrittstemperaturen um 1000°C, die heute schon technisch realisiert sind, sind chemische Reaktionen zur Erzeugung anderer Energieträger möglich. Beispielsweise kann Methan (CH_4) katalytisch gespalten werden. Die Spaltprodukte CO und H_2 lassen sich gut und verlustfrei transportieren und zur Energieerzeugung einfach wieder in Methan überführen. Viele mögliche chemische Reaktionen zur Energiespeicherung sind in Diskussion. Wasser läßt sich bei Temperaturen über 1100°C spalten. Die Energierückgewinnung ist auch besonders umweltfreundlich, da nun wieder Wasser entsteht.

5.3.6. Kernfusion

5.3.6.1. Fusionsreaktionen. Unter Fusion versteht man das Verschmelzen mehrerer (meist zweier) Kerne zu einem. Nach Abb. 5.2.3–1 sind Fusionen für leichte Kerne mit der Freisetzung von Energie verbunden. Besonders günstig ist die Fusion zu Helium $^4_2$He, da dieser Kern besonders stark gebunden ist und somit bei seiner Bildung besonders viel Energie an die Umgebung abgibt. Als Beispiel sei die Deuterium-Tritium-Reaktion angeführt: $^2_1\text{D} + ^3_1\text{T} \rightarrow ^4_2\text{He} + \text{n}$, bei der eine Energie von 17,6 MeV freigesetzt wird. Deuterium kann auch miteinander fusionieren, wobei 3,25 MeV Ener-

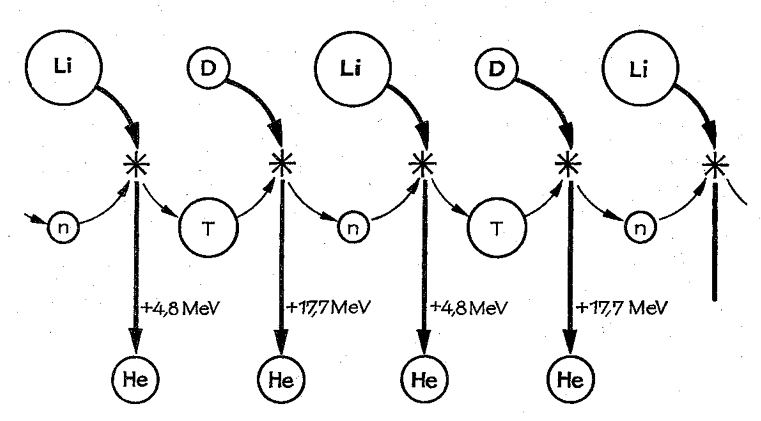


Abb. 5.3.6–1 Lithium-Deuterium-Fusionskette (T ≙ $^3_1$H, D ≙ $^2_1$H)

gie frei werden. Deuterium ist im normalen Wasser zu 0,015% statt Wasserstoff enthalten. Das klingt wenig, reicht aber für ewig aus, die Menschheit mit Energie zu versorgen. 1 Liter Wasser enthält durch dieses Deuterium eine Fusionsenergie, die dem Energieinhalt von 100 l Heizöl entspricht.

Tritium ist radioaktiv und hat eine Halbwertszeit von 12,3 Jahren. Es gehört somit nicht zu den natürlich vorkommenden Elementen, läßt sich aber im Reaktor aus dem ausreichend vorhandenen Lithium erzeugen: ${}^{6}_{3}\mathrm{Li} + \mathrm{n} \rightarrow {}^{3}_{1}\mathrm{T} + {}^{4}_{2}\mathrm{He} + 4{,}6\,\mathrm{MeV}$.

Fusionsbomben, die mit einem Li-D-Gemisch arbeiten, sind bekannt (Wasserstoffbombe), man spricht von der Lithium-Deuterium-Fusionskette, in der Lithium und Deuterium zu Helium fusionieren (Abb. 5.3.6–1). Die erste Fusionsbombe wurde vor Jahrzehnten gezündet, ein Fusionsreaktor wird erst in einigen Jahrzehnten zu realisieren sein.

Schwellwertenergien für Fusionsreaktionen. Außer der Anlagerung von Neutronen, erfordern alle Fusionen Energie zur Einleitung der Reaktion. Um Kerne so dicht aneinander zu bringen, daß sie über die Kernkräfte miteinander wechselwirken, ist erst einmal die Coloumb-Abstoßung der Protonen zu überwinden. Die Fusionsreaktionen leichter Kerne haben Reaktionsschwellen $\gtrsim 0{,}1$ MeV. Schon mit sehr einfachen Beschleunigern ist es leicht, Teilchen mit derartigen Energien zu erzeugen. In der experimentellen Kernphysik werden 14-MeV-Neutronen mittels der D-T-Fusionsreaktion durch Beschuß eines mit Tritium angereicherten Targets mit Deuteronen erzeugt. Das ist problemlos. Derart sind aber keine ausreichend hohen Teilchenströme erzeugbar. Für die Erzeugung von 1 MWh Wärmeenergie benötigt man mehr als 10^{18} Fusionen in der Sekunde. Alle laufenden Projekte für Fusionsreaktionen zur Energiegewinnung versuchen die Energieschwelle in einer Gasreaktion durch hohe Temperaturen zu überwinden.

Prinzipiell, wenn auch derzeit ohne Realisierungschancen für eine technische Verwertung, ist noch eine Möglichkeit in Diskussion, diese Energieschwelle zu umgehen. Man spricht von der müonkatalysierten Kernfusion.

Das Müon ist ein auch negativ geladen existierendes Elementarteilchen mit einer Lebensdauer von $2{,}2 \cdot 10^{-6}$ s. Es kann den Platz eines Elektrons im Atom einnehmen. Nukleonen und Müonen zusammen können ein „Atom" bilden *(Müonatom)*, was die bemerkenswerte Eigenschaft hat, wesentlich leichter als das entsprechende normale Atom mit anderen Atomen zu fusionieren. Das Müon ist ≈ 200mal schwerer als ein Elektron. Daraus ergibt sich ein Radius für die erste Bohrsche Bahn des Müonatoms von nur $2{,}5 \cdot 10^{-13}$ m ($0{,}5 \cdot 10^{-10}$ m im normalen Wasserstoffatom). Dieses sehr kleine neutrale Müonatom dringt leicht durch die Elektronenhülle anderer Atome und trägt seinen Kern sehr dicht an andere Kerne heran, ohne daß die abstoßenden Kern-Coulomb-Kräfte zur Wirkung kommen. Das vergrößert die Tunnelwahrscheinlichkeit beispielsweise bei der Verschmelzung von „Müondeuterium" mit Tritium. Auch die Bildung von Müonmolekülen und die anschließende Fusion ihrer Kerne ist wahrscheinlich. Allerdings ist die Lebensdauer der Müonen und damit auch die der Müonatome und -moleküle sehr gering. Im Gegensatz zu thermonuklearen Reaktionen sind diese Verschmelzungen „kalte" Fusionen. Bei der Kernfusion können wieder Müonen entstehen, Müonatome gebildet werden, die Müonen die Rolle eines Katalysators übernehmen.

Thermonukleare Reaktionen. In einem Gas sind Teilchenenergie und Temperatur verknüpft durch die Relation $\bar{W}_{\mathrm{kin}} = m/2\bar{v}^2 = 3/2kT$ mit $k = 8{,}6 \cdot 10^{-5}$ eV/K. Kernreaktionen, die über die Gastemperatur die notwendige Energie zur Auslösung der Reaktion erreichen, bezeichnet man als *thermonukleare Reaktionen.* Es sind sehr hohe Temperaturen notwendig. $W_{\mathrm{kin}} = 0{,}1$ MeV entsprechen Temperaturen von $7{,}7 \cdot 10^8$ K. Derartige Temperaturen sind bisher nicht beobachtet, die Temperatur im Inneren der Sonne entspricht nur Teilchenenergien von ≈ 2 keV. Die in der Sonne realisierte Fusion weist also auf Möglichkeiten, sie mit geringen Energien auszulösen. Hierfür spielen zwei Effekte eine Rolle. Zum einen errechnet sich über die Gastemperatur nur die mittlere Energie, es existieren aber auch, wenn auch in geringer Menge, Teilchen mit wesentlich höherer Energie. Zum anderen führt der quantenmechanische Tunneleffekt (vgl. 4.4.1.) mitunter schon zur Auslösung der Reaktion bei wesentlich geringerer Energie. Teilchen mit geringerer als der Schwellwertenergie, haben eine gewisse Wahrscheinlichkeit, den Coulomb-Wall zu durchtunneln und so in den Bereich der Kernkräfte zu kommen.

Lawson-Kriterium. Berücksichtigt man alle notwendigen Faktoren, eine thermonukleare Fusion einzuleiten, ergibt sich das Lawson-Kriterium: In Abhängigkeit von der verwendeten Gastemperatur sind für das Produkt aus Teilchendichte n und Einschlußzeit τ Mindestwerte zu erbringen, z.B. $40 \cdot 10^6$ K: $n \cdot \tau 10^{20}\,\mathrm{m}^{-3}\,\mathrm{s}^{-1}$.

Das hocherhitze Gas (Plasma) muß gegen seinen Expansionsdruck lange genug zusammengehalten werden, damit die Fusion erfolgen kann. Das Produkt $n \cdot \tau$ ergibt ab dieser Größe die Sicherheit, daß genügend Teilchen mit ausreichender Energie miteinander wechselwirken, etwa ab $n \cdot \tau = 3 \cdot 10^{20}\mathrm{m}^{-3}\,\mathrm{s}^{-1}$ brennt das Plasma ohne äußere Energieeinspeisung, durch Eigenerwärmung aus der Fusionsenergie selbständig weiter. Die notwendigen Temperaturen sind so hoch, daß keine Gefäßwandung zur Halterung des Plasmas verwendet werden kann – sie würde selbst zu Plasma zerfallen. Experimentell werden im wesentlichen zwei Wege beschritten, das Kriterium zu erfüllen.

1. *Trägheitshalterung.* Verdampft man eine (kleine) Materialmenge durch extrem schnelles Erhitzen, so dauert es einige Zeit, bis die Teilchen auseinander fliegen. Ihre Massenträgheit hält sie, wenn auch für eine sehr kurze Zeit, noch zusammen, so daß die Fusion erfolgen kann. In derartigen „Miniaturbomben" wird das Lawson-Kriterium für große n und kleine τ erfüllt.

2. *Magnetische Halterung*: Das Gasplasma kann mittels eines Magnetfeldes geführt und zusammengehalten werden. Die erreichbaren Plasmadichten n sind dabei nicht sehr hoch, was aber durch die prinzipielle Möglichkeit, große τ-Werte zu erreichen – stationärer Betrieb –, ausgeglichen wird.

5.3.6.2. Magnetische Halterung des Fusionsplasmas. Der auseinandertreibende Gasdruck des Plasmas ist in einfacher Näherung proportional dem Produkt aus der Teilchendichte und der Temperatur T, $P_{\mathrm{PL}} = n_{\mathrm{PL}} \cdot kT$, wobei unter der Teilchendichte n_{PL} die Dichte der positiv bzw. negativ geladenen Teilchen (n_+ bzw. n_-) zu verstehen ist (Plasma: $n_+ = n_- \approx 1/2n$), k = Boltzmann-Konstante.

Der aus der Elektrotechnik bekannte Pinch-Effekt erzeugt eine dem Teilchendruck entgegenwirkende Kraft. Bildet das Plasma in Form eines Kreisringes die Sekundärwicklung eines Transformators, können die Plasma-Ladungsträger kreisförmig beschleunigt werden und einen Stromfluß der Stärke I realisieren. Die Kraft des Pinch-Effektes erzeugt einen dem Gasdruck entgegenwirkenden Druck

$P_{\text{magn.}} \sim I^2 r^{-2}$. Der Druck ist proportional dem Quadrat der Stromstärke I und nimmt proportional mit dem Quadrat der Entfernung r vom Plasmafaden ab. Das Plasma haltert sich selbst bei $P_{\text{magn.}} = P_{\text{PL}}$.

Tokamak-Anlagen. Die fast ausschließlich für Kernfusionsanlagen verwendeten Experimentalanordnungen mit magnetischer Halterung sind Tokamak's. Das Prinzip wurde vor ≈ 30 Jahren von sowjetischen Physikern entwickelt. Sie prägten auch den Begriff *Tokamak*, entstanden aus den russischen Wörtern tok (Strom), kamera (Kammer) und magnitnige katuschi (Magnetspulen). Das Prinzip dieser Anordnungen ist in Abb. 5.3.6–2 dargestellt. Das Plasma befindet sich in einer vorher auf Ultrahochvakuum ausgepumpten Ringkammer. Das Plasma selbst hat eine Dichte von $n \lesssim 10^{20}\,\text{m}^{-3}$, was noch fünf Größenordnungen unterhalb der der Luft liegt. Die Ringkammer ist von Wicklungen zur Erzeugung eines das Plasma einschließenden toroidalen Magnetfeldes umgeben.

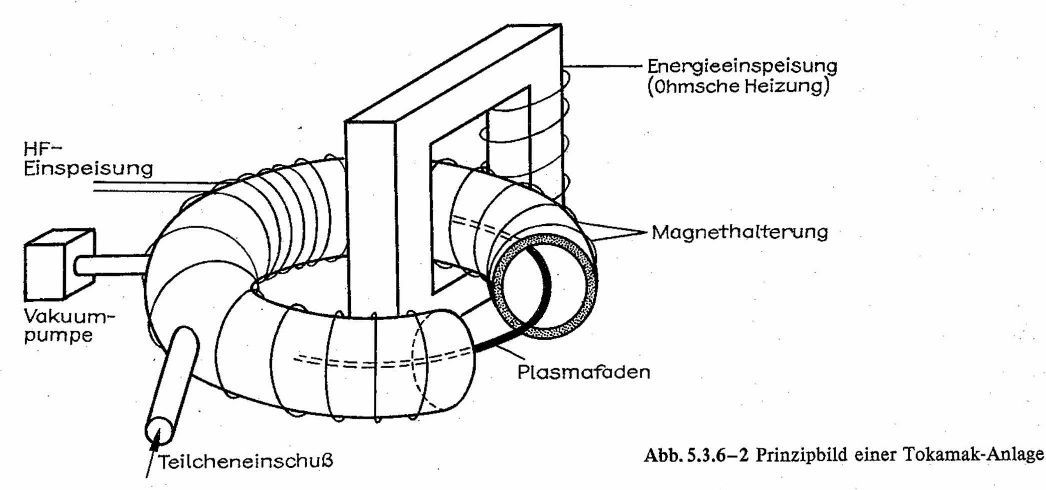


Abb. 5.3.6–2 Prinzipbild einer Tokamak-Anlage

Zusammen mit dem durch den Plasmastrom aufgebauten Magnetfeld, das den Pinch-Effekt erzeugt, ist man in der Lage, einen ringförmigen Plasmafaden zu stabilisieren. Bei den derzeit erreichbaren Magnetfeldstärken allerdings nur bis zu Teilchendichten von $n \sim 10^{21}\,\text{m}^{-3}$.

Über die Transformatoranordnung werden die Plasmateilchen beschleunigt, wird ihnen Energie zugeführt. Die „Ohmsche" Heizung ist aber prinzipiell nicht in der Lage, ausreichend hohe Plasmatemperaturen zu erzeugen. Zusätzlich muß noch durch die Einspeisung über Hochfrequenzsender und durch den Einschuß von stark beschleunigten Neutralteilchen dem Plasma weitere Energie zugeführt werden. Die experimentellen Anlagen sind mittlerweile derart aufwendig geworden und die technologischen Probleme derart vielgestaltig, daß man versucht, weltweit für einen zukünftigen Fusionsreaktor Kosten und Forschungskapazitäten gemeinsam aufzubringen.

Dieses INTOR-Projekt sieht vor, einen gewaltigen Tokamak zu errichten, in dem ein Plasma mit $n \cdot \tau = 1{,}3 \cdot 10^{22}\,\text{m}^{-3}\text{s}$ bei 10^8 K 100 s lang stabilisiert werden kann. Die beteiligten Länder UdSSR, USA, Japan und die der Euratom planen, etwa 1990 diesen vor Fusionskraftwerksanlagen letzten Experimentalreaktor mit schon merklicher Abgabe von Fusionsenergie in Betrieb zu nehmen.

Derzeit glaubt man, alle wesentlichen Probleme der stabilen Halterung des Plasmas auch bei höchsten Temperaturen gelöst zu haben. Eine wirtschaftliche Variante sieht man in der Entwicklung von *Fusionshybridreaktoren*, das sind Fusionsanlagen, bei denen die freiwerdenden schnellen Neutronen noch zum Erbrüten von Spaltstoffen (Plutonium) genutzt werden.

5.3.6.3. Trägheitshalterung des Fusionsplasmas. Die zweite Hauptvariante zur Entwicklung von Fusionskraftwerken sieht die Erfüllung des Lawson-Kriteriums, im Gegensatz zur magnetischen Halterung, für hohe Teilchendichten und kurze Einschließungszeiten vor. Wenn man in der Lage ist, ein Material innerhalb von $\lesssim 10^{-9}$ s auf Fusionstemperatur zu bringen, wird es aufgrund seiner Massenträgheit noch kurzzeitig verharren, ehe es auseinander treibt. Auf diesem Verharrungsprinzip beruht die Trägheitshalterung. Die am weitesten vorangetriebene Variante der Trägheitshalterung verwendet Laser zur Aufheizung. In Laserfusionsanlagen wird das in kleinen Kügelchen eingeschlossene Deuterium-Tritium-Brennstoffgemisch im freien Fall in einer Vakuumkammer allseitig von Laserimpulsen getroffen, aufgeheizt und explodiert wie eine Mikrowasserstoffbombe.

Die Brennstoffkugel in der Größe um 0,1 mm Durchmesser enthält das Deuterium-Tritium-Brennstoffgemisch auf bis zu 10^8 Pa komprimiert. Die nur 1/1000 mm dicke Umhüllung kann aus vielen Schichten verschiedener Materialien bestehen. Sie soll mit der eingestrahlten Laserleistung durch ihren Rückstoß beim Verdampfen das Fusionsgemisch stark komprimieren. Es werden Dichten, die das Zehnfache der von Festkörpern entsprechen, erreicht, so daß sich mit den realisierten Aufheizzeiten um 10^{-9} s die Bedingung zur Fusion erfüllen läßt. Die erforderlichen Laserleistungen werden durch die Entladung großer Kondensator-Batterien erreicht; die in der kurzen Zeitspanne abgegebene Energie entspricht der Leistung aller Kraftwerke Europas. Es werden die Strahlen vieler Höchstleistungslaser aus allen Richtungen auf das Kügelchen fokussiert. Die Gleichmäßigkeit der Bestrahlung seiner Oberfläche ist auf 1 % genau notwendig. Das erfordert die derzeit höchstmögliche feinmechanische Präzision mit elektronischer Einstellkontrolle und Steuerung. Derzeit werden in Laser-Fusionsanlagen schon Fusionen mit einer Ausbeute von über 10^{10} Neutronen realisiert. Es laufen auch Projekte, bei denen die

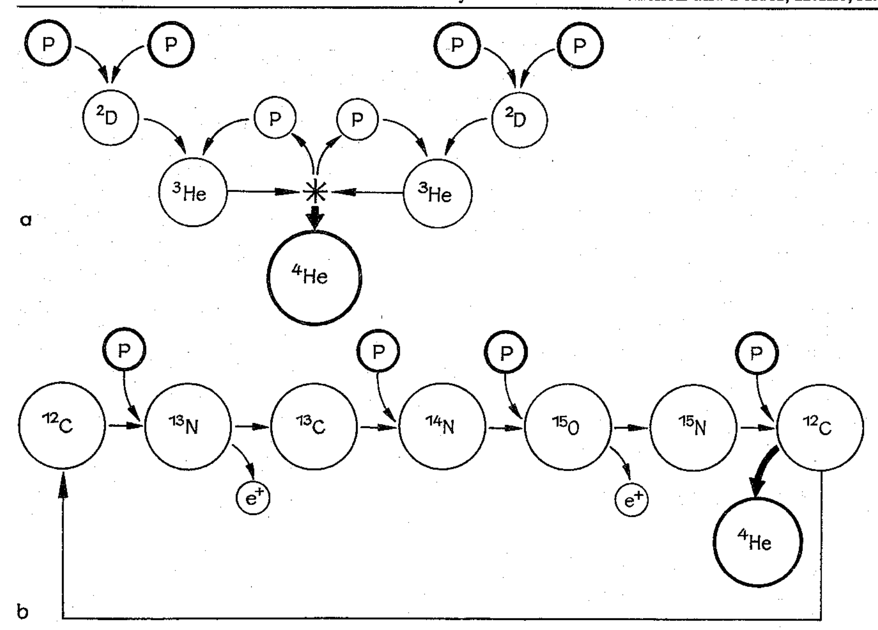


Abb. 5.3.6–3 (a) Proton-Proton-Fusionskette, (b) Kohlenstoff-Stickstoff-(Bethe-Weizsäcker-)Zyklus

Energieeinspeisung mit Elektronen- bzw. Ionenstrahlen erfolgt. Gegenwärtig kann noch nicht entschieden werden, welche Aufheizvariante am günstigsten ist, auch ist noch nicht entscheidbar, ob Fusionsanlagen mit Trägheitshalterung – etwa Laserfusionsanlagen – oder die mit magnetischer Halterung nach dem Tokamak-Prinzip die zu realisierende Fusionskraftwerksvariante sein wird.

5.3.6.4. Fusion in den Sternen. Die Energieerzeugung in den Sternen geschieht durch Fusion. In Abhängigkeit von der herrschenden Temperatur sind verschiedene Fusionsreaktionen möglich und wahrscheinlich. Ab Temperaturen von $T \gtrsim 10^8$ K kann Helium zu schwereren Elementen verschmelzen. Bei der Temperatur unserer Sonne von $\approx 1{,}5 \cdot 10^7$ K (im Zentrum) sind Fusionen zu Helium wahrscheinlich. Betrachtet man das Lawson-Kriterium, existieren in der Sonne bei einer Dichte von $n \sim 10^{30}$ Teilchen/m³ und unendlicher Einschließungszeit ($\tau = \infty$) Idealbedingungen zur Fusion, zumal die Sonne zu 73 Masseprozenten aus Wasserstoff besteht. Zwei wesentliche Reaktionszyklen, die vom Wasserstoff zu Helium führen, werden angenommen. Beide haben die Bilanzgleichung $4\,^1_1\text{H} \rightarrow {}^4_2\text{He} + 2e^+ + 2\nu + 4 \cdot 10^{-12}$ J. Man rechnet damit, daß derart je Sekunde $6 \cdot 10^{11}$ kg (600 Mill. Tonnen) Wasserstoff verbraucht werden und die Sonnenmasse von $2 \cdot 10^{30}$ kg noch ausreicht, über 10^{11} Jahre unvermindert weiter Energie zu spenden. Die beiden Reaktionszyklen, Proton-Proton-Kette und Kohlenstoff-Stickstoff-Zyklus sind in Abb. 5.3.6–3 dargestellt. Der C-N-Zyklus, auch als *Bethe-Weizsäcker-Zyklus* bezeichnet, benötigt den Kohlenstoff nur als Katalysator.

5.3.7. Kernwaffen

Als *Kernwaffen* bezeichnet man für den militärischen Einsatz vorgesehene Sprengkörper. Die Wirkung wird durch die bei der Spaltung oder Fusion freiwerdende Energie erzielt. Es ist üblich, abgesehen von den Strahlenwirkungen aller Art, die Sprengkraft mit dem stärksten konventionellen Sprengmittel, dem 2,4,6-Trinitrotoluen (TNT) zu vergleichen. Der Vergleich führt in die Bereiche von Kilo- und Millionen-Tonnen TNT-Äquivalent. Die Spaltenenergie von 1 g U-235 ($\triangleq 8{,}2 \cdot 10^{10}$ J) entspricht der Explosionsenergie von $2 \cdot 10^4$ kg TNT.

5.3.7.1. Wirkungsprinzip nuklearer Sprengsätze. Eine Zerfallsmöglichkeit für die instabilen Nuklide U-235 und Pu-239 ist die Kernspaltung. Dieser natürliche Vorgang kann durch den Beschuß mit Neutronen erzwungen werden. Da bei der Spaltung zugleich auch freie Neutronen entstehen, können diese weitere Spaltungen auslösen und so eine Kettenreaktion in Gang setzen (vgl. 5.3.4.). Die Kettenreaktion läuft bei einer ausreichenden Menge spaltbaren Materials an. Die kritische Masse beträgt bei U-235 ≈ 50 kg. In einer Mischung mit Moderatormaterial, umgeben von einem Neutronenreflektor, ist diese kritische Masse um zwei Größenordnungen reduzierbar. Der Reflektor kann über die Reaktion ⁹Be (n, 2n)⁸Be auch als Neutronenvervielfacher arbeiten. Die Kettenreaktion läuft sehr schnell ab, in $\approx 10^{-6}$ s wäre die gesamte kritische Masse spaltbaren Materials aufgebraucht, wenn es nicht ihr entgegenwirkende Faktoren gäbe. Durch die freiwerdende Energie verdampft auch das noch spaltbare und alles sonstige Material der Umgebung, so daß ohne besondere Halterung keine hohe Ausbeute erreicht wird. Das ist aber nicht wesentlich; die über Hiroshima abgeworfene Bombe hatte auch nur ≈ 1 kg U-235 (entspricht $2 \cdot 10^7$ kg TNT) in Energie umgesetzt. Die Masse der Gesamtkonstruktion war so bemessen, daß sie von einem Flugzeug noch transportiert werden konnte. Eine drei Wochen zuvor gezündete Bombe von etwa halber Sprengkraft wog über $2 \cdot 10^5$ kg. Die weitere Entwicklung ging in Richtung der Sprengkrafterhöhung bei Verringerung des Gesamtgewichtes. Der als *Atombombe* be-

zeichnete Kernspaltungssprengkörper wurde 1952 durch die Fusionsbombe (*H-* oder *Wasserstoffbombe*) noch übertroffen. In der Sprengkraft wurde die Größenordnung von 10^{10} kg TNT-Äquivalent erreicht.

Kernfusionssprengkörper. Um eine Kernfusion in Gang zu setzen, sind bei hohem Druck sehr hohe Temperaturen notwendig (vgl. 5.3.6.). Kernspaltungssprengsätze eignen sich, mit der Erzeugung von Temperaturen in der Größenordnung von 10^7 K über eine Zeit von $\sim 10^{-6}$ s, als Zünder für die Lithium-Deuterium-Fusionsreaktion. Für die Fusion müssen Lithium und Deuterium in hoher Dichte vorhanden sein. Hierzu wird es verfestigt und bei der Zündung durch konventionelle Sprengmittel oder durch die Kernspaltungsreaktion stark komprimiert. Die thermonukleare Fusion beruht auf der thermischen Aufheizung der Reaktionspartner zur Überwindung der Zündbarriere. Durch die bei eingeleiteter Fusion freiwerdende Energie kann die Reaktionstemperatur aufrechterhalten werden. Während für Kernspaltungsbomben vor der Zündung unterkritische Mengen des Spaltstoffes getrennt angeordnet werden müssen, und dadurch nicht beliebig große Mengen zündbar sind, hängt die Größe der Sprengkraft einer Fusionsbombe nur von der Menge des verwendeten Materials ab Lithium und Deuterium sind in genügender Menge auf der Erde vorhanden und in reiner Form darstellbar. Die größten bekannten Wasserstoffbomben sind so bemessen, daß ihre Sprengkraft ausreicht, eine Großstadt mit Millionen Einwohnern total zu zerstören und alles Leben in ihr zu vernichten.

Kernminen und Neutronenbomben. Die weitere Entwicklung der Waffentechnik führte durch Fortschritte bei den Trägersystemen zu einer Verbesserung der Treffergenauigkeit, was eine Reduzierung der Sprengkraft erlaubte. Auch Konzeptionen, Kernwaffen ohne weltweite Eskalation nur im Frontbereich einzusetzen, führten zur Entwicklung immer kleinerer Kernsprengsätze. Da die Fusionssprengkörper Kernspaltungszünder benötigen, beruhen die Kernwaffen mit kleinstem Wirkungsradius ausschließlich auf Kernspaltung. Die kleinste derzeit zur Auslösung der Kettenreaktion notwendige Menge Pu-239 bzw. höchstangereichertes U-235 liegt bei $\approx 0{,}5$ bis 5 g entsprechend 10^4 bis 10^5 kg TNT. Die kleinsten Einheiten sind tragbar.

Die Kernspaltungs- oder Fusionsbomben begleitenden Strahlenwirkungen sind vielfältig. Unterhalb einer Sprenkraft von $5 \cdot 10^5$ kg TNT dominieren – pauschal abgeschätzt – die Strahlenschäden. Die Strahleneffekte lassen sich in gewissen Grenzen ohne Erhöhung der Sprengkraft verstärken. Eine starke Strahlenwirkung, bei einer Sprengkraft von $\approx 10^6$ kg TNT, entwickelt die *Neutronenbombe*. Die D-T-Fusionsreaktion setzt 80 % ihrer Energie in Form schneller Neutronen ($W_{\text{kin}}(n) = 14$ MeV) frei. Das erforderliche Tritium kann in Kernreaktoren aus Lithium erzeugt werden oder (mit geringerer Effektivität) über die gleiche Reaktion bei der Explosion selbst entstehen. Die biologisch hochwirksamen Neutronen (vgl. 5.4.6.) sind in Luft auf Reichweiten von 1 km beschränkt. Um die gleiche Wirkung in 2 km Entfernung zu erreichen, ist eine ≈ 350fache Kernladung notwendig. Optimiert man die Reichweite der tödlichen Neutronendosis bei minimaler mechanischer Zerstörung, ergibt sie sich zu 1 km Entfernung vom Detonationsort. Eine Explosion 500 m über dem Erdboden würde dabei keine größeren Zerstörungen anrichten. Notwendig sind ≈ 1 g D-T-Gemisch und ein Kernspaltungszünder mit einer Wirkung $\lesssim 4 \cdot 10^5$ kg TNT (20 g gespaltenes U-235).

Nuklearer elektromagnetischer Impuls (EMP). Ein Nebeneffekt bei Kernexplosionen im All, weniger auf der Erde selbst, eventuell nur von Interesse als physikalisches Phänomen, ist der EMP. Es ist ein erst in jüngster Zeit den Fachleuten im zivilen Sektor bekannt gewordener Effekt.

Wird im erdnahen Raum, beispielsweise in einigen 100 km Höhe, eine Kernwaffe gezündet, so entsteht mit einem Energieanteil von einigen Prozent auch elektromagnetische Strahlung. Diese Simultanstrahlung ist auf eine Zeitdauer $< 10^{-7}$ s beschränkt. Sie breitet sich geradlinig aus, und der die oberen Luftschichten erreichende Anteil ionisiert dort mittels Compton-Effekt, d. h. stößt Elektronen aus ihren Atomen, die sich mit hoher Energie in Strahlungsrichtung bewegen. Sie wirken nun wie ein Teilchenstrom, der, durch starke Sonnenaktivität ausgelöst, das bekannte Nordlicht erzeugt. Auch die Compton-Elektronen werden im erdmagnetischen Feld gekrümmt und dadurch Quelle eines intensiven elektromagnetischen Sekundärimpulses aller Frequenzen, einschließlich des Gebietes der Radiostrahlung. Dieser Impuls ist von noch etwa gleicher Kurzzeitigkeit und damit extrem stark. Es werden Feldstärken von $10^5\,\text{Vm}^{-1}$ angegeben. Damit wird im Sichtbereich der Explosion auf der Erde jegliches leitfähiges Material, als Antenne wirkend, einen starken elektrischen Impuls aufnehmen.

1962 wurde in 400 km Höhe eine 1,5-Megatonnen-(TNT-Äquivalent-)-Wasserstoffbombe gezündet. Im 1 500 km entfernten Honolulu ging die Straßenbeleuchtung aus, die Stromversorgung brach zusammen. Heutzutage würden alle Radios, Fernseher, fast jegliche Elektronik zerstört. So verlockend diese Möglichkeiten für Waffenentwickler sein mögen, so gut sind auch die Anlagen zumindest der Militärtechnik, gegen den EMP abzusichern. Die Komplikation ist nur, daß derart kurzzeitige Störungen bisher nicht zu bewältigen waren, da selbst Blitzschutzanlagen nur um Größenordnungen langsamer zu reagieren brauchen.

5.3.7.2. Art und Stärke der Wirkung nuklearer Sprengsätze. Bei Kernwaffen, abgesehen von denen mit verstärkter Strahlenwirkung, ergibt sich eine Aufteilung der Wirkung entsprechend ihrer Energieabgabe: Stoßwellenenergie 50 %, thermische Energie 35 %, simultane Strahlung 5 %, Folgestrahlung 10 %. Die Stoßwellenenergie ist in Tab. 5.3.7–1 in Abhängigkeit von der Sprengkraft als tödlicher Radius dargestellt. Die Detonationsstärke ist so gewaltig, daß bei 10^{10} kg TNT-Äquivalent es bis zu 8 km Entfernung zur vollständigen Zerstörung von Gebäuden kommt und in 15 km Entfernung noch mit schweren Gebäudeschäden zu rechnen ist. Die Wirkung der thermischen Strahlung führt bei 10^6 kg TNT-Äquivalent noch in 1 km Entfernung zu Verbrennungen zweiten Grades. Bei 10^7 kg kommt es noch in 10 km Entfernung vom Epizentrum durch eine Blendwirkung 100mal stärker als die Sonne zu Netzhautverbrennun-

Tab. 5.3.7–1 Tödlicher Radius einer Kernexplosion in Abhängigkeit von der Sprengkraft

Explosionskraft TNT-Äquivalent in kg	tödlicher Radius in km
10^6	0,2
10^7	0,5
10^8	1,5
10^9	3,0
10^{10}	6,6

gen, bei $2 \cdot 10^8$ kg reicht der Entflammungspunkt 4,5 km weit. Bei den schweren Kernwaffenladungen überwiegen in der sofortigen Wirkung Stoßwelle und thermische Strahlung die der radioaktiven Teilchen- oder γ-Strahlung bei weitem. Die schweren Ladungen hinterlassen eine starke Langzeitaktivität, die Gebiete auf Jahrzehnte unbewohnbar werden läßt.

Bei kleinen Kernwaffenladungen ($< 5 \cdot 10^5$ kg TNT-Äquivalent) überwiegen die Wirkungen der γ- und der Neutronenstrahlung. Bei Ladungen ohne verstärkte Strahlenwirkung reicht die Grenze der tödlichen Strahlung für 10^6 kg TNT-Äquivalent über 0,5 km weit, für die Neutronenwaffe gleicher Stärke bis $\approx 1{,}2$ km.

Tab. 5.3.7–2 zeigt Daten aus amerikanischer Literatur nach Versuchen an Affen über die Wirkung schneller Neutronen, wie sie die Neutronenbombe abstrahlt. Bis 500 m sind die betroffenen Personen noch etwa bis zu 5 min handlungsfähig und werden in einer Entfernung von ≈ 700 m innerhalb von zwei Stunden funktionell gestört.

Schutzwirkungen sind nur im Sinne des Erhaltens der Handlungsfähigkeit für einen Gegenschlag von Wert. Es existiert keine Möglichkeit eines umfassenden Schutzes auch nur der Zivilbevölkerung gegen schwere Kernwaffen. Verschiedene Materialien haben verschiedene Abschirmwirkungen für die verschiedenen Strahlenarten. Von der Teilchenstrahlung ist nur die der schnellen Neutronen aus der Spaltung und insbesondere der D-T-Fusion von Bedeutung. Wasser und Beton schwächen die Neutronenintensität je 30 cm Schichtdicke um eine Größenordnung. 90 cm Beton absorbieren danach 99,9 % der Neutronen und sind schon ein wirksamer Schutz, zumal die Neutronen ihre Energie, ohne wesentliche Aktivität zu erzeugen, abgeben. Für Panzer sind verschiedene Materialien und Materialkombinationen (bevorzugt mit Bor) entwickelt, die die Neutronenintensität bis auf 10 % abschirmen können.

γ-Strahlung hat eine bis zu fünfmal geringere biologische Wirksamkeit. Die Intensitäten sind derart, daß die gleiche Strahlenbelastung etwa mit einer um eine Größenordnung stärkeren Sprengkraft erreicht wird.

Die Unmöglichkeit eines umfassenden Schutzes menschlichen Lebens vor Kernwaffenwirkungen ergibt sich aus der enormen möglichen Sprengkraft, die jeden Schutz durchbrechen kann und der möglichen Langzeitaktivierung der gesamten Umwelt bis zur Unbewohnbarkeit. Hinzu kommen mögliche Klimaveränderungen durch Staubanreicherung, die als *nuklearer Winter* bezeichnet werden.

5.4. Experimentelle Methoden der Kern- und Elementarteilchenphysik

5.4.1. Quellen energiereicher Teilchen und γ-Quanten

5.4.1.1. Elektronen- und Ionenquellen. Ladungsträger verschiedenen Vorzeichens sind über ihre elektrostatische Anziehung bestrebt, sich zu kompensieren. Daher sind neutrale Systeme vorherrschend, und man muß sich bei Bedarf Ladungsträger erst erzeugen. Zur Erzeugung freier Elektronen nutzt man überwiegend den Prozeß der thermischen Emission, in speziellen Fällen ist die Feldemission von Vorteil, andere Elementarprozesse spielen als intensitätsstarke Elektronenquellen keine Rolle. Um Elektronen abzulösen, ist in Gasen die Ionisierungsenergie aufzubringen und bei Festkörpern die Austrittsarbeit zu leisten. Die Austrittsarbeit beträgt z. B. für Wolfram 4,54 eV, für thoriertes Wolfram 2,63 eV und für Bariumoxid auf Platinnickel 1,0 eV.

Feldemission erfolgt bei hohen elektrischen Feldstärken aus Festkörperoberflächen. Leicht an Spitzen zu erzeugende Feldstärken von $10^9\,\mathrm{Vm^{-1}}$ können durch Absenken des Potentialtopfrandes die Austrittsarbeit um mehr als 1 eV verringern. Die Emission läßt sich durch den quantenmechanischen Tunneleffekt beschreiben. Die Feldemissionsstromdichte nimmt ab einer bestimmten Feldstärke exponentiell mit der Feldstärke zu. Durch die Feldemission entstehen relativ monoenergetische Elektronen, die z. B. in der hochauflösenden Elektronenmikroskopie von Bedeutung sind. Feldemissionselektroden erfordern Ultrahochvakuum, da schon monoatomare Bedeckungen die Potentialverhältnisse wesentlich ändern.

Tab. 5.3.7–2 Wirkung einer Kernladung mit verstärkter Strahlungskraft von 10^6 kg TNT-Äquivalent

Entfernung in m	Dosis in Gy	Zeit bis der Tod eintritt in Tagen
400	180	1 bis 2
500	80	
640	30	5 bis 7
760	6,5	

Thermo- oder Glühemission. Hierunter versteht man das Auslösen von Elektronen aus direkt oder indirekt geheizten Materialien. Das Verdampfen der Elektronen aus Metallen (Glühkatoden) ist die gebräuchlichste Methode zur Elektronenstrahlerzeugung. Die Elektronenstromdichte j wird durch die Richardson-Gleichung beschrieben, $j = a \cdot T^2 \exp(-W_A/kT)$, ($W_A$ = Austrittsarbeit, k = Boltzmannkonstante, T = Temperatur, kT = thermische Energie der Elektronen). a ist eine Konstante, die für reine Metalle $\approx 6 \cdot 10^5\,\mathrm{A\,m^{-2}\,K^{-2}}$ beträgt.

Ionenquellen erzeugen überwiegend positiv geladene Ionen durch Stoßionisation aus der Gasphase. (Die Erzeugung negativ geladener Ionen erfolgt durch Elektronenanlagerung an Atome und Moleküle). Von Bedeutung ist aber auch die Ionenemission aus Festkörperoberflächen durch Zerstäuben mit energiereichen Teilchen und die thermische Ionenemission. Bei letzterem Effekt dampfen Teilchen von einer genügend geheizten Unterlage als Ionen ab, wenn ihre Ionisierungsenergie geringer als die Austrittsarbeit der Atome der Unterlage ist. Derart kann man die verschiedensten Ionen, meist aus Oxiden mit Wolframpulver verpreßt und erhitzt, erzeugen. Diese Stromquellen liefern aber nur Stromdichten von 10^{-10} bis $10^{-8}\,\mathrm{A/m^2}$.

Ionenquellen existieren in den verschiedensten Ausführungen, z. B. als Vakuumfunken-, Hochfrequenz-, Niedervoltbogen-, Elektronenstoß- oder Kanalstrahlrohrionenquelle. Elektronenstoßionenquellen haben den in Abb. 5.4.1–1 dargestellten prinzipiellen Aufbau. Diese Finkelstein-Ionenquelle besteht aus einem Ionisierungsraum, in dem ein Druck von $\approx 10^{-6}$ Pa des zu ionisierenden Gases (z. B. Wasserstoff, Argon, C_6H_6), aufrechterhalten wird. Die ionisierenden Elektronen entstehen an der Glühkatode und werden zum Anodenzylinder hin beschleunigt. Die Ionisierungswahrscheinlichkeit erhöht man durch Pendelbahnen der Elektronen zwischen verschiedenen Elektroden, oder wie in der Abbildung dargestellt, durch Spiralbahnen mittels eines Magnetfeldes. Die Ionisation ist bei ≈ 3- bis 5fa-

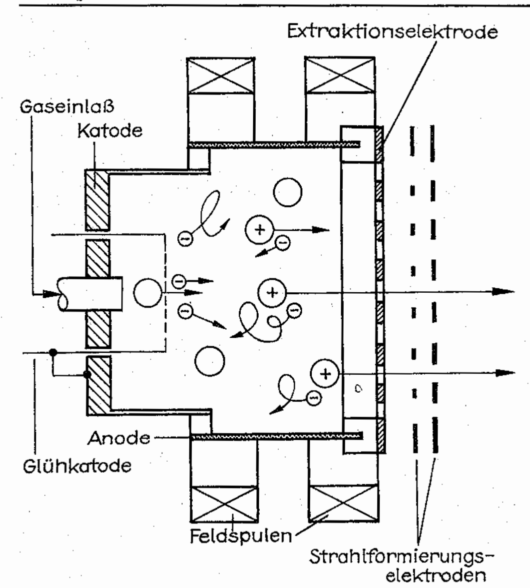


Abb. 5.4.1–1 Elektronenstoßionenquelle

cher Ionisierungsenergie am effektivsten. Die Ionisierungsenergien betragen z. B. 13,6 eV für Wasserstoff, 12,1 eV für Xenon und 24,6 eV für Helium. Zur Erzeugung von Ionenströmen im Ampere-Bereich sind Bogenentladungsionenquellen entwickelt. Bei z. B. 10^{-4} Pa brennt bei diesen zwischen einer kräftigen Glühkatode und einer oft als Lochblende ausgebildeten Anode ein Plasma hoher Teilchendichte. Eine auf Hochspannung liegende Extraktionselektrode reicht mit ihrem Feld durch die Anode durch bis in das Plasma und saugt die Ionen ab.

5.4.1.2. Strahlerzeugungssysteme. Elektronen- und Ionenströme sind bequem durch elektrische und magnetische Felder formbar. In den meisten Fällen benötigt man nicht nur die geladenen Teilchen überhaupt, sondern Teilchen definierter Art und Energie. Etwa mit homogener Flächenverteilung oder als punktförmige Quelle zur elektronen- bzw. ionenoptischen Weiterverarbeitung z. B. in Beschleunigern. Das Extraktionssystem ist zumeist schon Teil der Strahlerzeugungsanordnung. In Abb. 5.4.1–1 folgen der Extraktionselektrode zwei weitere Elektroden zur Strahlformierung. In vielen Fällen werden heute für selbst großflächige Bestrahlungen auf ±1 % über die Strahlfläche homogene Teilchenströme gefordert. Für fokussierende Anlagen sind Ionenquellen kleinster Apertur und höchster Stromdichte erwünscht. Inwieweit die Quelle punktförmig ist, wird durch die engste Stelle des Ionenstrahles bestimmt; der von dort aus vom Strahl gebildete halbe Öffnungswinkel ist die Apertur des Systems. Die Elemente zur Elektronen- und Ionenstrahlformierung und -führung sind an anderer Stelle behandelt (vgl. 2.6.4.1.).

5.4.1.3. Neutronenquellen. Neutronen sind nicht mit elektrischen oder magnetischen Feldern ablenkbar und auch nur unter Intensitätsverlust zu reflektieren (spiegeln). Man kann einen Strahl durch ein Loch in einer dicken Abschirmwand ausblenden (Kollimator). Gebräuchlichste *Neutronenquelle* ist der Kernreaktor. Entweder gelangen durch Reaktorkanäle Neutronen in den Außenraum zu Experimenten, oder aber – was meist für hohe Neutronenflußdichten der Fall ist – die zu bestrahlenden Proben gelangen über diese Kanäle an spezielle Orte im Reaktor. Im Reaktor können dafür Zonen mit überwiegend schnellen Neutronen (bis 10 MeV) ausgewählt werden, oder Bereiche, wo die thermischen Neutronen überwiegen. Neutronenflußdichten von $10^{19}\,m^{-2}\,s^{-1}$ werden erreicht. An speziellen Kernreaktoren, die im Impulsbetrieb arbeiten, kann eine zeitauflösende Spektroskopie betrieben werden. Auch oft genutzte Neutronenquellen entstehen mittels Kernreaktionen an Teilchenbeschleunigern. Für schnelle Neutronen eignen sich u. a. die Reaktionen ${}^2_1D + {}^2_1D \rightarrow {}^3_2He + n$, ${}^3_1T + {}^2_1D \rightarrow {}^4_2He + n$. Für langsame Neutronen werden Schwellwertreaktionen verwendet, wobei für die Reaktion ${}^3_1T + {}^1_1H \rightarrow {}^3_2He + n$ die Wasserstoffatome mit einer Energie knapp oberhalb des Schwellwertes von 1,02 MeV auf das mit Tritium angereicherte Target geschossen werden. Transportable Neutronenquellen ermöglicht die Reaktion ${}^9_4Be + \alpha \rightarrow {}^{12}_6C + n$, wobei α-Teilchen aus radioaktiven Präparaten verwendet werden. Ein Radium-Beryllium-Gemisch ist eine viel genutzte Neutronenquelle mit kontinuierlicher Energieverteilung bis 13,7 MeV.

5.4.1.4. Erzeugung von γ-Quanten. Als γ-Quanten bezeichnet man elektromagnetische Strahlung mit kürzeren Wellenlängen und höheren Energien als Röntgenstrahlen. Berühmt geworden ist insbesondere für Anwendungen in der Medizin die *Cobalt-Quelle*. Durch Neutronenbestrahlung entsteht aus dem in der Natur vorkommenden Co-59 das langlebige Isomer Co-60. Es zerfällt zum größten Teil mit einer Halbwertszeit von 5,24 Jahren in einen angeregten Zustand des ^{60}Ni, der unter Aussendung zweier γ-Quanten in den Grundzustand übergeht (Abb. 5.4.1–2). Die γ-Quanten haben Energien von 1,17 und 1,33 MeV. In der Medizin eingesetzte Quellen haben Aktivitäten von $\sim 10^{13}$ Bq (1 Becquerel ≡ 1 Umwandlung/s). Die Strahlungsintensität entspricht der von ≈ 300 g Radium.

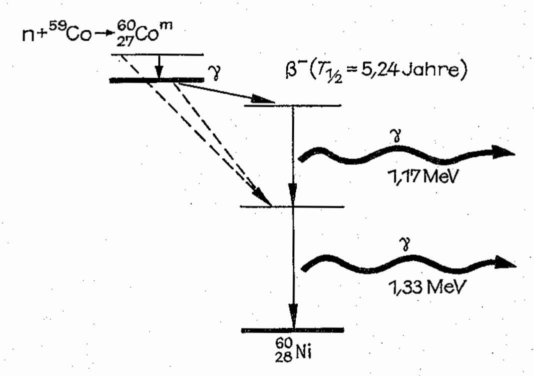


Abb. 5.4.1–2 Zerfall von Cobalt-60

Auch Elektronenbeschleuniger, z. B. Betatrons, werden zur Erzeugung von γ-Quanten eingesetzt. Die γ-Quanten entstehen wie in der Röntgen-Röhre (vgl. 5.6.5.) durch Abbremsung der Elektronen im Coulomb-Feld der Atome eines Targets. Die Elektronen eines Betatrons von beispielsweise 50 MeV werden auf ein intensiv gekühltes Wolframtarget gelenkt, das dann Quelle der Bremsstrahlung von bis zu 50-MeV-Energie ist. Natürlich sind radioaktive Elemente selbst auch Quelle energiereicher γ-Quanten, da die meisten radioaktiven Zerfälle von γ-Strahlung begleitet sind.

Synchrotronstrahlung. Für den Wellenlängenbereich unterhalb des Sichtbaren bis in das Gebiet der γ-Strahlen gibt es seit den 60er Jahren eine Strahlenquelle besonders guter Qualität. Seit MAXWELLS Elektrodynamik ist bekannt, daß beschleunigt bewegte elektrisch geladene Teilchen elektromagnetische Wellen abstrahlen. Die Strahlung beim Abbremsen der Elektronen in Röntgen-Röhren war dafür bisher das klassische Beispiel. Neuerdings beginnen spezielle Beschleuniger der Kernphysik in einigen Einsatzfällen der Röntgen-Bremsstrahlung Konkurrenz zu machen. Die an Synchrotronen zuerst gefundene Strahlung von im Kreise umlaufenden Elektronen (Synchrotronstrahlung) ist zu einer sehr begehrten Quelle energiereicher Strahlung – zumindest für die wissenschaftliche Analytik – geworden. Es sind spezielle Beschleuniger und Speicherringe allein zur Nutzung dieser Strahlung entwickelt und gebaut worden.

Auf gekrümmten Bahnen laufende Elektronen haben auch bei dem Betrage nach konstanter Geschwindigkeit eine Komponente der Beschleunigung (sie bewirkt die Zentrifugalkraft). Die Frequenz des abgestrahlten Lichtes entspricht im nichtrelativistischen Falle bei Kreisbahnen der Umlauffrequenz. Bei annähernd Lichtgeschwindigkeit liegt sie um Größenordnungen höher. Die derzeitigen Beschleuniger für Synchrotronstrahlung erreichen Wellenlängen bis unterhalb 10^{-10} m. Die Intensität der Strahlung ist proportional der vierten Potenz der Energie der umlaufenden Teilchen, sie liegt um Größenordnungen über der konventioneller Strahlungsquellen. Zudem sind in den Ringbeschleunigern besondere Ablenkeinheiten eingebaut (Wigglermagnete, Undulatoren), mit denen die Intensität der Strahlung noch verfielfacht werden kann. Synchrotronstrahlung ist polarisiert, der elektrische Vektor liegt in der Bahnebene. Da in den Beschleunigern die Elektronen gebündelt umlaufen, ist die Synchrotronstrahlung gepulst. Die Experimente sind darauf einzurichten, daß in den Impulsen eine sehr hohe Strahlungsdichte relativ zur integralen Dichte vorliegt.

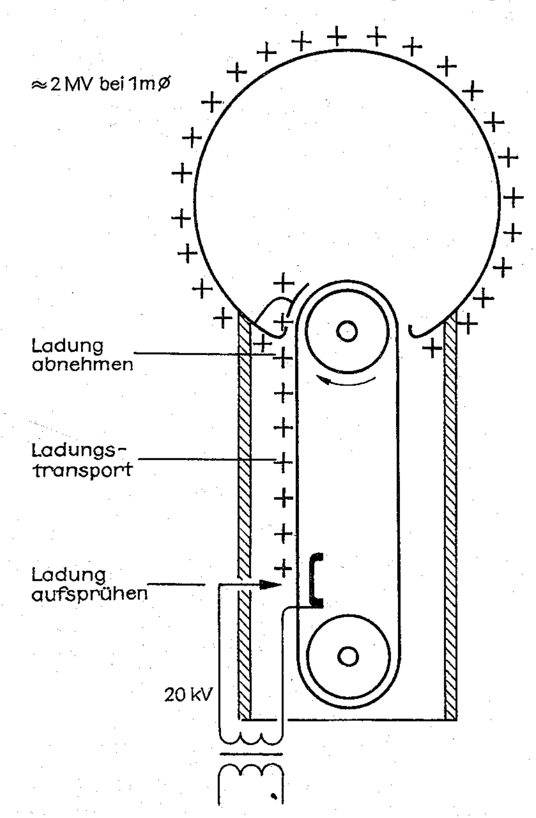


Abb. 5.4.2–1 Band- oder Van-de-Graaff-Generator

5.4.2. Linearbeschleuniger

5.4.2.1. Kaskaden- und Bandgeräte. *Beschleuniger* sind allgemein Geräte, in denen geladene Teilchen beschleunigt werden, in denen ihnen kinetische Energie übertragen wird. Im elektrischen Feld ist die aufgenommene kinetische Energie gleich dem Produkt von Teilchenladung q und durchlaufender Spannung U, $W_{\text{kin}} = q \cdot U$.

Um ein einfach geladenes Teilchen (z. B. ein Elektron) auf eine Energie von $W_{\text{kin}} = 25$ keV zu beschleunigen (Fernsehröhre), muß es eine Spannung von 25 kV durchlaufen (Spannung zwischen Katode und Anode). Ein radioaktives Präparat liefert α-Teilchen mit einer Energie von 4,8 MeV. Um zweifach ionisierte Heliumatome auf diese Energie zu beschleunigen, müssen sie ein Potential von 2,4 MV durchlaufen.

Die Erzeugung hoher Spannungen ist Grundvoraussetzung zum Betrieb von Beschleunigern. In Einfachbeschleunigern ist das durchlaufene Potential gleich der angelegten Spannung U, in anderen Beschleunigern wird ein Potential mehrfach durchlaufen und die kinetische Energie ist dementsprechend ein Mehrfaches des Produktes $q \cdot U$.

Hohe Spannungen erreicht man beispielsweise über einen Hochspannungstransformator und anschließender Spannungsvervielfachung. Ohne besondere Isolationsschwierigkeiten sind Spannungen auf ≈ 100 kV transformierbar. Durch eine spezielle Schaltung von Dioden und Kondensatoren (Grainacher-Schaltung; Kaskadengenerator) erreicht man Gleichspannungen von einigen Megavolt. An Luft sind derart Spannungen bis ≈ 2 MV isolierbar.

Bandgeneratoren (*van-de-Graaff-Generatoren*) sind auch geeignet, Spannungen bis zur Grenze der Isolationsfähigkeit des Mediums der Umgebung (Luft oder Schutzgas) zu erzeugen (Abb. 5.4.2–1). Der Generator nutzt den Effekt, daß das Innere eines Leiters feldfrei ist. Einer Hohlkugel kann man innen so lange Ladungen zuführen, bis diese so starke Felder erzeugen, daß die Ladungen außen absprühen. Die Kapazität einer Kugel ist durch ihren Radius gegeben und beträgt an der Luft für 30 cm Durchmesser etwa 600 kV und für 1 m Durchmesser ≈ 2 MV. Das Aufbringen der Ladungen erfolgt über ein isolierendes Transportband mit Bandgeschwindigkeiten von 60 m s^{-1}, wobei Ströme bis zu einigen Milliampere transportiert werden können. Die aufzusprühende Ladungsmenge wird einem Hochspannungstransformator entnommen. Um die Grenzspannung heraufzusetzen, werden die Bandgeneratoren in Drucktanks eingebaut. Ein N_2-CO_2-Inertgasgemisch von 1 bis 1,5 MPa gestattet es, Spannungen bis 12 MV zu isolieren. Van-de-Graaff-Generatoren zeichnen sich durch eine gute Energiehomogenität und -stabilität aus. Um geladene Teilchen Höchstspannungen durchlaufen lassen zu können und den überschlagfreien Betrieb bei Grenzfeldstärken von bis zu 30 MV m^{-1} (in Drucktanks) zu gewährleisten, sind besondere Anordnungen entwickelt worden.

Beschleunigungsstrecken bestehen aus einer Kette von im

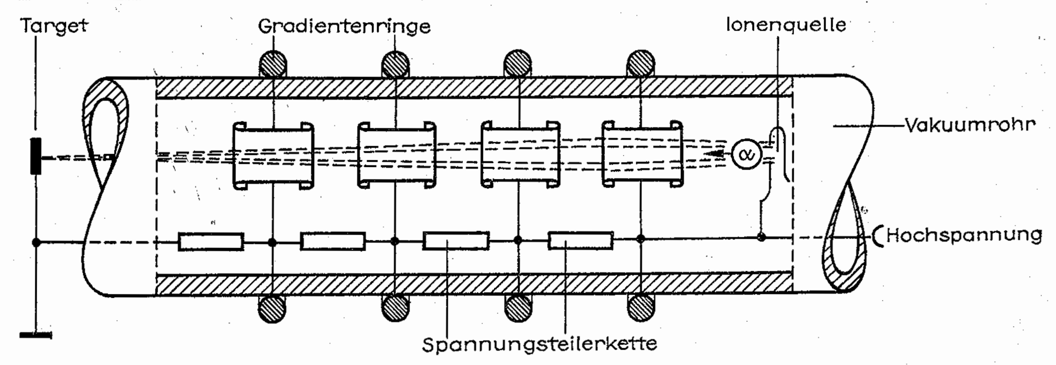


Abb. 5.4.2–2 Beschleunigungsstrecke

Hochvakuum angeordneten Zylinderlinsen, durch die die zu beschleunigenden Teilchen schrittweise das Gesamtpotential durchlaufen und als Strahl formiert und geführt werden (Abb. 5.4.2–2). Sogenannte Gradientenringe homogenisieren den Feldverlauf und werden, wie die Linsen, über eine Spannungsteilerkette auf das zugehörige Potential gebracht. Es ist praktisch, die Ionenquelle auf Hochspannung zu legen, um am Target – im Experimentierraum – auf Erdpotential arbeiten zu können. Eine Beschleunigungsstrecke bildet zusammen mit der Hochspannungserzeugung den linearen Einfachbeschleuniger. Einfachbeschleuniger (Van-de-Graaff- und Kaskadenbeschleuniger) haben den Nachteil, nur Teilchen mit einer Energie $W_{kin} = q \cdot U$ erzeugen zu können. Die Energie ist durch die maximal erzeugbare Spannung U begrenzt. Diese Beschleuniger liefern aber im Gegensatz zu allen im weiteren zu besprechenden Anlagen einen Teilchengleichstrom und haben sich z. B. zum Implantieren in der Halbleiterindustrie oder zum Auslösen entsprechender Kernreaktionen für Neutronengeneratoren bewährt.

Tandembeschleuniger verdoppeln die Teilchenhöchstenergie der Einfachbeschleuniger. Sie bestehen aus der Stromerzeugung des Einfachbeschleunigers und aus zwei kompletten hintereinander geschalteten Beschleunigungsstrekken. In der ersten werden negative Ionen auf maximale Energie beschleunigt. Danach durchlaufen die Ionen einen Gasraum oder durchdringen „Stripper"-Folien, wobei sie zum Teil umionisiert werden und als positive Teilchen das Potential in dem zweiten Beschleunigerrohr – noch einmal den gleichen Energiebetrag gewinnend – mit umgekehrter Polung durchlaufen. Die Beschleunigung erfolgt von Erdpotential auf Erdpotential. Die Umladung ist nicht sehr effektiv, weshalb eine mehrfache Nutzung des Effektes in der Praxis nicht realisiert ist.

5.4.2.2. Lineare Hochfrequenzbeschleuniger (Vielfachbeschleuniger). In der in Abb. 5.4.2–3 gegebenen prinzipiellen Anordnung lassen sich kurze „Teilchenbündel" durch vielfaches Durchlaufen *einer* Spannungsdifferenz prinzipiell beliebig hoch beschleunigen. Driftrohrlänge, Abstände und die Frequenz sind so abgestimmt, daß ein diese Beschleunigungsstrecke durchfliegendes Teilchen oder Teilchenbündel immer gerade im Feld zwischen den Rohren eine beschleunigende Feldrichtung durchläuft. Die umgekehrte Polung der Wechselspannung kommt nicht zur Wirkung, da dann das Teilchen oder Teilchenbündel sich im feldfreien Raum innerhalb der Driftrohre befindet. Die Synchronisierbedingung verlangt sich stetig vergrößernde Driftrohrlängen a_i, die durch die Geschwindigkeit der Teilchen und der Periodendauer T gegeben sind, zu $a_i = v_i \cdot T/2$, $U = U_0 \sin(2\pi/T \cdot t)$.

Die Beschleuniger arbeiten im Dezimeterwellen-Bereich und erreichen Teilchenenergien bis 15 MeV m^{-1}. Die Anzahl der Driftrohre ist nur durch den technischen Aufwand begrenzt. Ionenbeschleuniger mit ≈ 100 MeV werden bevorzugt zur Einspeisung vorbeschleunigter Teilchenbündel in große Kreisbeschleuniger eingesetzt. Besonders günstig sind lineare Hochfrequenzbeschleuniger für Elektronen herstellbar. Bei einer Frequenz von 3 GHz kann nach 0,3 m

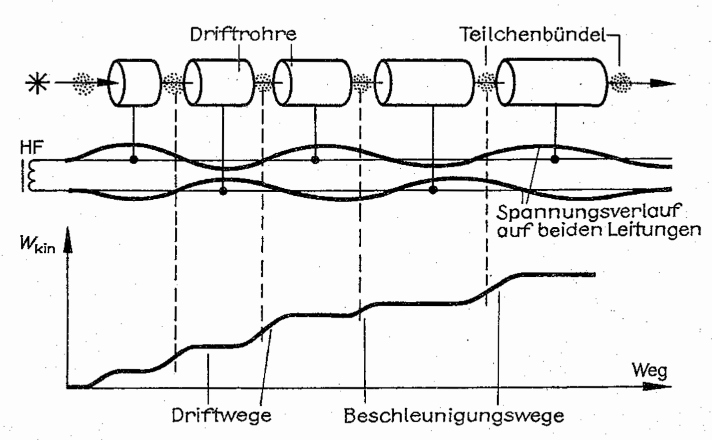


Abb. 5.4.2–3 Hochfrequenzbeschleunigerstrecke

die zehnfache Ruheenergie der Elektronen erreicht werden. Sie haben dann annähernd Lichtgeschwindigkeit, so daß zur Weiterbeschleunigung nur noch konstante Driftrohrlängen (10 cm) benötigt werden. Haben die Teilchen schon nahezu Lichtgeschwindigkeit, können sie nach einem anderen Prinzip effektiver auf noch höhere Energien beschleunigt werden.

Wanderwellenbeschleuniger arbeiten nach dem Wellenreiterprinzip. Wenn sich Teilchen und elektrisches Feld in einem Hohlleiter mit gleicher Geschwindigkeit ausbreiten, kann das Teilchen in der immer beschleunigenden Phase mitlaufen (auf der abfallenden Flanke der Wechselspannungswelle reiten) und dem Feld Energie entziehen. Die Phasengeschwindigkeit der Hochfrequenz wird durch besondere Formgebung für die Wellenleiter eingestellt. Derartige Beschleuniger haben noch den Effekt der Selbstsynchronisation (Autophasierung). Bleiben Teilchen hinter der Phasengeschwindigkeit etwas zurück, gelangen sie in einen steileren Potentialgradienten und werden stärker beschleunigt, eilen sie der Phase voraus, kommen sie in ein Gebiet mit flacherem Potentialgradienten und werden etwas weniger beschleunigt. Für Forschungszwecke sind gewaltige Hochfrequenzbeschleuniger entstanden, z. B. in Stanford (USA) ein 20-GeV-Beschleuniger von 3 km Länge.

5.4.3. Kreisbeschleuniger

Allen Kreisbeschleunigern ist die Verwendung von Magnetfeldern zur Krümmung der Teilchenbahnen eigen. Das Magnetische Führungsfeld B bewirkt über die Lorenz-Kraft $F_L = q \cdot v \cdot B$ (vBF; q = Teilchenladung, v = Teilchengeschwindigkeit) eine Bahnkrümmung. Der Bahnradius R stellt sich über das Gleichgewicht der Kräfte – Lorenz-Kraft ist gleich Zentrifugalkraft $F_Z = mv^2 R^{-1}$ – ein, $R = mv/(q \cdot B)$.

Je größer der Impuls mv eines beschleunigten Teilchens ist, umso größer muß das das Teilchen auf einer Kreisbahn mit dem Radius R haltende Magnetfeld B sein.

5.4.3.1. Zyklotrone. Alle Zyklotronvarianten haben als Teilchenbahnen Spiralen. Die einfachste Form, das *klassische Zyklotron*, arbeitet mit konstantem Magnetfeld. Abb. 5.4.3–1 zeigt das Funktionsprinzip. Das magnetische Führungsfeld bewirkt die Bahnkrümmung. Zwischen den metallischen Halbschalen D („Dees") werden die Teilchengruppen in einem Hochfrequenzfeld beschleunigt und spiralen in immer größer werdenden Halbkreisen Energie gewinnend nach außen. Ihre Umlauffrequenz ist unabhängig vom Radius, weil Bahnlänge und -geschwindigkeit gleichermaßen zunehmen. Die Frequenz der Beschleunigungsspannung wird auf die Umlauffrequenz $\nu = q \cdot H/2\pi m$ abgestimmt. Im Bereich nichtrelativistischer Teilchenenergien kann mit konstanter Frequenz gearbeitet werden. Im Bereich relativistischer Geschwindigkeiten wird die in der Formel für die Umlauffrequenz enthaltene Masse m geschwindigkeitsabhängig und die Teilchenpakete kommen bei fester Frequenz außer Takt. Aus diesem Grund sind Festfrequenzzyklotrone zur Beschleunigung von Elektronen ungeeignet. Maximale Energien ergeben sich bei Bahnradien von etwa 1 m zu ≈ 10 MeV für Protonen und 40 MeV für α-Teilchen.

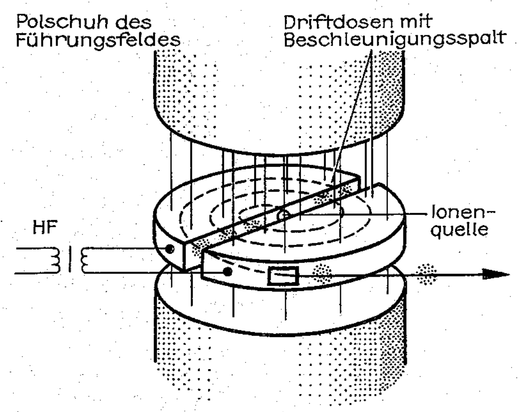


Abb. 5.4.3–1 Prinzipskizze des Zyklotrons

Synchrozyklotron. Läßt man nur ein Teilchenpaket im Zyklotron umlaufen, kann die Beschleunigungsfrequenz der im relativistischen Falle geschwindigkeitsabhängigen Umlauffrequenz des Teilchenpaketes angeglichen, synchronisiert werden. Ein derart arbeitendes Gerät bezeichnet man als Synchronzyklotron (frequenzmoduliertes Zyklotron). Das berühmteste Gerät dieser Art steht in Dubna bei Moskau. Der Polschuhdurchmesser des 7000-t-Magnetes beträgt 6 m. Mit einer Hochfrequenz von 300 kV Scheitelspannung lassen sich Protonen auf 680-MeV-Energie beschleunigen.

Isochronzyklotrone halten die Umlauffrequenz der Teilchenpakete durch eine Magnetfeldänderung konstant, so daß auch für relativistische Energien noch mit fester Frequenz beschleunigt werden kann. Das Magnetfeld wird durch besondere Polschuhformung mit wachsendem Radius dementsprechend ansteigend gestaltet. Mit supraleitenden Magnetspulen können die Abmessungen der Zyklotrone verringert werden. Zyklotrone liefern Teilchenströme bis ≈ 100 µA und sind dadurch trotz ihrer begrenzten Maximalenergie vorteilhaft einzusetzen.

Mikrotron (Elektronenzyklotron). Die Umlaufzeit τ in einem Zyklotron mit der Magnetfeldstärke H für ein Teilchen der Ladung q ist: $\tau = 2\pi m_0/(qH)\,[1 + W_{kin}/(m_0 c^2)]$. Sobald die Bahnenergie W_{kin} der Teilchen nicht mehr gegen die Ruheenergie m_0c^2 zu vernachlässigen ist, wird die Umlaufzeit energieabhängig und ist nicht mehr im Takt der Frequenz der beschleunigenden Hochfrequenz. Zyklotrone sind allgemein zur Beschleunigung von Elektronen nicht geeignet, da deren Ruheenergie nur $m_{e,0}c^2 = 0{,}511$ MeV beträgt. Die Ruheenergie ist aber so gering, daß sie in einem Beschleunigungsakt als Energiezuwachs realisiert werden kann. Durchlaufen Elektronen eine Potentialdifferenz von 511 kV, nimmt ihre Energie um m_0c^2 zu und ihre Umlaufzeit ist dann gerade ein ganzzahliges Vielfaches der Umlaufzeit im nichtrelativistischen Falle; sie sind wieder im Takt mit der Hochfrequenz.

Ein derart arbeitendes Elektronenzyklotron wird auch als Mikrotron bezeichnet. Strahlungsverluste der mit relativ starker Krümmung umlaufenden Elektronen bewirken eine Grenzenergie von ≈ 10 MeV.

5.4.3.2. Betatron (Elektronenschleuder). Das *Betatron* ist ein Kreisringbeschleuniger, in dem die Energiezuführung durch ein sich zeitlich änderndes magnetisches Feld erfolgt. Es arbeitet nach dem Transformatorprinzip, wobei die Sekundärentwicklung aus den im Kreise umlaufenden Elektronen besteht. Abb. 5.4.3–2 veranschaulicht die Funktionsweise. Die Primärwicklung erzeugt im Kern des Magneten ein sich entsprechend der angelegten Wechselspannung änderndes Magnetfeld, das die in dem Vakuumrohr

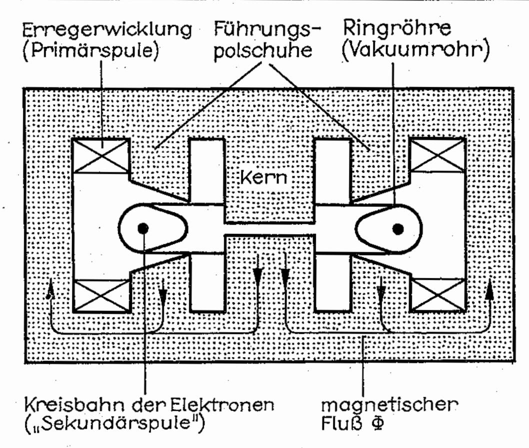


Abb. 5.4.3–2 Schematischer Schnitt durch ein Betatron

umlaufenden Elektronen beschleunigt. Die Elektronen würden sofort ihren Sollkreis verlassen, wenn nicht ein sich gleichzeitig mit änderndes Führungsfeld die Elektronen dort stabilisieren würde. Die Bedingung zur Beschleunigung auf einem Sollkreis ist erfüllt, wenn die Magnetkerne so aufgebaut sind, daß das magnetische Führungsfeld $\boldsymbol{B}_F$ sich stets so mit dem beschleunigenden Kernfeld $\boldsymbol{B}_K$ mit ändert, daß immer $\boldsymbol{B}_F = \boldsymbol{B}_K/2$ ist. Eine an die Primärwicklung angelegte Wechselspannung von z.B. 50 Hz erfüllt diese Bedingung im linearen Teil des Anstieges. In der Zeit von ≈ 3 ms werden die Elektronen über $\approx 10^6$ Umläufe auf Energien von einigen 10 MeV beschleunigt. Problematisch ist das zeitgenaue Einschießen der Teilchen in die Kreisbahn und das Auslenken nach der Beschleunigung. Die Teilchen mit ausreichend hoher Anfangsenergie und genau tangential einzuschießen, gelingt nur mit Elektronen. Bei ≈ 400 MeV Elektronenenergie würde die zugeführte Energie durch die Strahlungsdämpfung aufgebracht werden – ist die theoretische Grenzenergie erreicht. In der Praxis haben sich Betatrons bis 50 MeV durchgesetzt. Sie dienen in der Mehrzahl der Erzeugung von γ-Strahlen für medizinische und technische Zwecke.

5.4.3.3. Synchrotrone. Allen Zyklotronvarianten ist die für die spiralartigen Bahnen notwendigerweise große Polschuhfläche eigen. Der Vorteil von Beschleunigern, bei denen Teilchen auf einem Sollkreis laufen, liegt in den geringeren Aufwendungen für die das Führungsfeld erzeugenden Magnete. Synchrotrone sind Kreisringbeschleuniger mit einem geregelten Führungsfeld, um das zu beschleunigende Teilchenpaket auf dem Sollkreis zu haltern. Die zur Beschleunigung angelegte Hochfrequenz ist mit der Geschwindigkeit des Teilchenpaketes synchronisiert. Synchrotrone sind zumeist Großanlagen. Erste Synchrotrone besaßen noch gewaltige Ringröhren mit den Abmessungen eines Autotunnels. Erst als man die Teilchenbahnen genauer einregeln konnte, sank der Querschnitt unter 100 cm². Das Synchrotron in Serpuchow bei Moskau hat als Beispiel bei einem Sollkreisdurchmesser von 470 m einen Ringröhrenquerschnitt von $0{,}07 \times 0{,}14$ m². Die sehr gute Fokussierung der Teilchen auf dem Sollkreis wird durch eine spezielle Magnetfeldanordnung erreicht, bei der abwechselnd über den Kreis verteilt horizontal und vertikal fokussiert wird. Dieses Magnetfeld mit alternierendem Gradienten führte zur Bezeichnung *AG-Synchrotron*. Synchrotrone werden für Elektronen und Nukleonen getrennt entwickelt. Prinzipiell ist der Aufbau in beiden Fällen der gleiche. Die Grenzenergie der Elektronensynchrotrone ist trotz der geringen Krümmung bei den großen Sollkreisradien durch die Abstrahlung (Synchrotronstrahlung) gegeben, sie liegt bei ≈ 20 GeV (20 000 MeV).

Das derzeit größte Elektronensynchrotron in Hamburg beschleunigt Elektronen auf 6 GeV. Auf dem Beschleunigungsring von 100 m Durchmesser sind abwechselnd Beschleunigungsstrecken und Magneteinheiten angeordnet. Die Elektronen werden in einem Linearbeschleuniger auf ≈ 40 MeV vorbeschleunigt und dann tangential in den Kreis eingeschossen.

Protonensynchrotrone mit Radien bis über 1 km liefern Energien um 500 GeV ($5 \cdot 10^{11}$ eV). Beschleuniger mit supraleitenden Magnetspulen zur Erhöhung der maximalen Feldstärke und weiter verringertem Röhrenquerschnitt für Energien von mehreren Teralektronenvolt (10^{12} eV) sind im Bau. Die großen Beschleuniger tragen klangvolle Namen: Synchrophasotron, Bevatron (engl., 1 BeV = 1 GeV), Cosmotron (erreicht Energien der kosmischen Strahlung).

5.4.3.4. Speicherringe. Mit wachsender Endenergie der Nukleonen aus großen Beschleunigern ergab sich ein Problem. Stößt bei kleinen Energien ein beschleunigtes Proton ein Targetproton, so steht maximal die halbe kinetische Energie des stoßenden Teilchens für Nukleonenwechselwirkungen zur Verfügung, die andere Hälfte geht ungenutzt als kinetische Energie beider Stoßpartner „verloren". Das ist noch eine gute Energieausnutzung. Bei kinetischen Energien, die nicht mehr gegen die Ruheenergie der Stoßpartner zu vernachlässigen sind, wird die nutzbare Energie $W_{\text{nutz}} = 2W_0 \quad [(1 + W_{\text{kin}}/2W_0)^{1/2} - 1]$, $\quad W_0 = m_0c^2$; $\quad W_0$ (Proton) = 0,938 GeV. Für den Stoß zweier Protonen ist die Energienutzung in Tab. 5.4.3–3 für einige W_{kin}-Werte angegeben. Sie weist aus, daß oberhalb 10 GeV der Aufwand zur Erzeugung höherer Energien in derartiger Experimentierweise kaum noch zu rechtfertigen ist. Ein Ausweg aus dieser Situation ist gegeben, wenn nicht mehr ein bewegtes gegen ein ruhendes Teilchen geschossen wird, sondern beide Teilchen gegeneinander beschleunigt aufeinander treffen.

Tab. 5.4.3–3 Nutzbare Energie beim Stoß eines Nukleons mit der kinetischen Energie W_{kin} gegen ein ruhendes Nukleon

W_{kin} in GeV	W_{nutz} in GeV	Prozent
2,5	1	40
6	2	30
70	10	14
5 200	100	2

Bei gleichschweren Teilchen, die mit gleichen kinetischen Energien genau aufeinander treffen, steht die gesamte kinetische Energie als Reaktionsenergie zur Verfügung. Eine Möglichkeit, das zu realisieren, wären zwei gegeneinander arbeitende Beschleuniger. Eine billigere Variante ermöglichen *Speicherringe*. Das sind „Parkräume" für beschleunigte Teilchenpakete. Abb. 5.4.3–4 zeigt den prinzipiellen Aufbau von experimentellen Anlagen mit Speicherringen. Ein Kreisringbeschleuniger speist Teilchenpakete in zwei übereinander angeordnete Speicherringe. Die Speicherringe haben auch wie die Beschleuniger Durchmesser in der Grö-

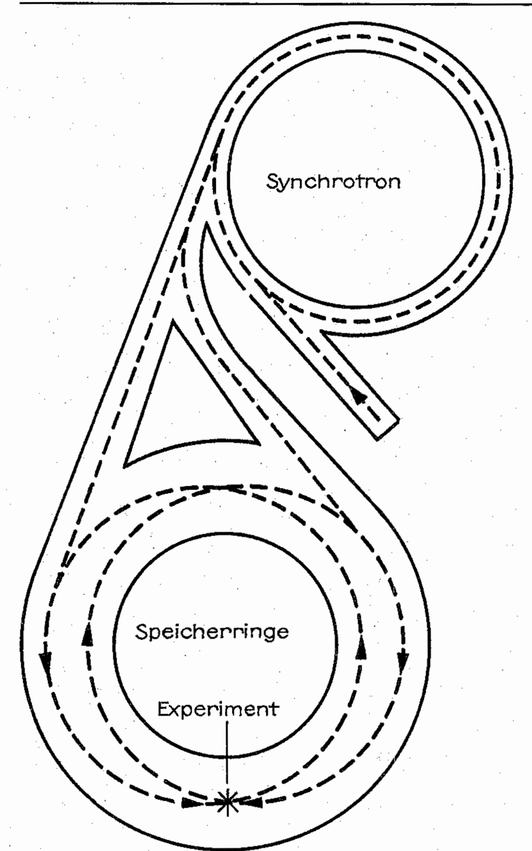


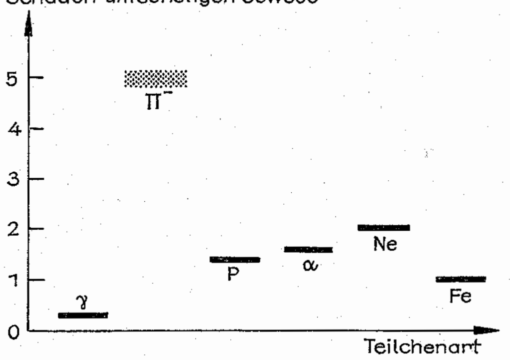
Abb. 5.4.3–4 Experimentieranlage mit Speicherringen

ßenordnung 100 m. Abwechselnd angeordnete Magnet- und Beschleunigersegmente halten die Teilchen auf ihrem Sollkreis bzw. kompensieren die Strahlungsverluste. In den Ringröhren herrscht ein derart gutes Ultrahochvakuum, daß der Strahl 24 Stunden zu halten ist. Die Speicherringe werden nach und nach mit Teilchenströmen von bis zu 1 A aufgefüllt, die dann an dafür vorgesehenen Stellen gegeneinander gelenkt werden können.

5.4.3.5. Einsatz von Beschleunigern für technische Zwecke. Die jeweils größten Beschleuniger der Welt werden ausschließlich für Probleme der Grundlagenforschung konzipiert und eingesetzt. Aber was vor zwei bis drei Jahrzehnten auf diesem Gebiet das Neuste war, wird heute mitunter in größeren Stückzahlen z. B. in der Technik oder Medizin nutzbringend verwendet.

Stromstarke Linearbeschleuniger für Ionen von einigen 10^5 eV werden in der modernen Halbleitertechnologie eingesetzt. Die Ionen dringen z. B. als Dotierungselement in das Halbleitermaterial Silicium in eine durch die Energie festlegbare Tiefe ein. Sie können dann in einer gewissen Breite um die mittlere Reichweite z. B. eine elektrische Sperrschicht bilden. Mitunter sind Beschleuniger in medizinischen Zentren zur Produktion kurzlebiger Isotope für diagnostische Zwecke eingesetzt.

Betatrons werden überwiegend als Quellen harter γ-Strahlen genutzt. Die Behandlung von Krebsgeschwülsten erfolgt wesentlich durch Bestrahlung. Mitunter werden dadurch operative Eingriffe unnötig oder überhaupt erst möglich durch Rückbildung von Schwellungen. Problematisch ist bei der Behandlung von nicht an der Oberfläche liegenden Geschwülsten die unumgängliche Schädigung gesunder Gewebepartien. Zwar wird schon auf den Behandlungsort im Körper fokussiert und das dazwischenliegende Gewebe durch Strahlverbreiterung so gut es geht geschont, aber dennoch belastet. Besonders gut für den Einsatz in der Medizin scheinen π-Mesonen geeignet zu sein, was aus Abb. 5.4.3–5 ersichtlich ist. Aufgetragen ist im oberen Teil ein Modellbeispiel für die relative Schädigung des zu zerstörenden Gewebes im Vergleich zur Schädigung des gesunden Gewebes der Umgebung. Die γ-Strahlung entstammt dem Co-60-Isotop. Im unteren Teil der Abbildung findet man die Erklärung für die Wirkungsweise des π-Mesons. Es dringt entsprechend seiner Geschwindigkeit und des Abstandes vom Entstehungsort (in diesem Beispiel 15 cm tief) in das Gewebe ein, ohne wesentliche Wechselwirkung. Dann zerfällt es am Ende seiner Lebensdauer von $2{,}6 \cdot 10^{-8}$ s mit starker biologischer Wirksamkeit. Das entspricht fast einer Punktwirkung am gewollten Ort. Mesonenstrahlen können z. B. mit größeren intensitätsstarken Zyklotronen über Kernreaktionen erzeugt werden.



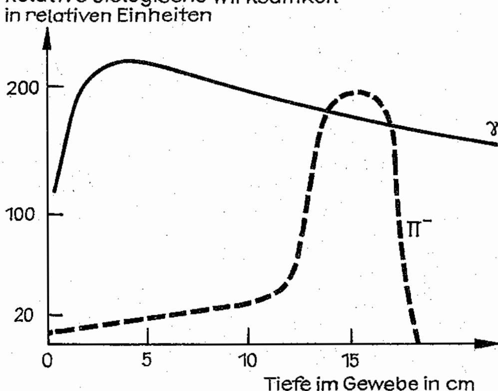


Abb. 5.4.3–5 Wirkung verschiedener Teilchenstrahlen und von γ-Quanten des Co-60 bei der Neutralisierung einer Geschwulst

Elektronenbeschleuniger, das klassische Betatron und zunehmend auch das mit ≈ 1 m Durchmesser kleine Mikrotron werden im Erzbergbau eingesetzt. Sie liefern Elektronen von z. B. 10 MeV Energie, mit denen im Gestein charakteristische Röntgen-Strahlung angeregt werden kann. Mit ihr lassen sich vorhandene Elemente, z. B. Kupfer, leicht bestimmen. Die quantitative Analyse gestattet bei kleiner Probenmenge noch den Nachweis von 0,1 g Gold auf 1 t Erz.

5.4.4. Nachweisgeräte für elementare Teilchen

Einfachste Form des Nachweises elementarer Teilchen, wie γ-Quanten, Elektronen, Protonen oder α-Teilchen, sind die Schwärzung einer Fotoplatte oder das im Dunkeln zu beobachtende Aufblitzen in einer Leuchtschirmschicht nach Auflegen eines radioaktiven Präparates.

5.4.4.1. Szintillationszähler. Verschiedene Materialien szintillieren (Aufblitzen, Aufleuchten) unter der Wirkung elementarer Teilchen. In den Anfängen der physikalischen Meßtechnik mußte man sich damit begnügen, für quantitative Auswertungen, mit der allerdings im Dunkeln enormen Empfindlichkeit des menschlichen Auges, die Zahl der meist sehr schwachen Lichtblitze je Zeiteinheit zu bestimmen.

Abb. 5.4.4–1 zeigt das Prinzip des gebräuchlichsten Szintillationszählers. Die Lichtblitze des NaI-Kristalles lösen in der auf der Rückseite aufgebrachten Fotokatode Elektronen aus, die in einem *Sekundärelektronenvervielfacher* zu einem kräftigen Stromimpuls anwachsen. Sekundärelektronenvervielfacher (SEV) sind Vakuumröhren mit mehreren hintereinander liegenden Prallelektroden, die so angeordnet und mit Spannung versorgt werden, daß die Elektronen von einer Elektrode zur anderen reflektieren und bis zur letzten, der Auffangelektrode beschleunigt werden. Da bei jeder Reflexion ein Elektron immer noch mehrere neue herausschlägt, vervielfacht sich die Anzahl um den Faktor 10^6 bis 10^8, wodurch ein kräftiger, leicht meßbarer Stromimpuls am SEV abgenommen werden kann.

Für Elektronen ist kein Szintillator notwendig, kann die erste Elektrode direkt als Detektorfläche dienen. Mitunter läßt man noch das Eintrittsfenster weg (offener SEV), wenn die Energieverluste darin zu groß sind, kann dann aber den Vervielfacher nur in Vakuumapparaturen einsetzen. Moderne Bauformen bestehen nur noch aus dünnen (z. B. 0,1 mm ∅) Spezialglasröhren (Channeltrons). Die Elektronen werden an der Innenwand reflektiert, vervielfachen sich, und dadurch, daß das Glas schwach leitend ist, entsteht bei an den Enden angelegter Spannung das zur Beschleunigung der Elektronen notwendige elektrische Feld.

Szintillationszähler sind neben der reinen Zählung ionisierender Teilchen auch zur Messung der kinetischen Energie einsetzbar, da die Zahl der Fotoelektronen und damit die Höhe der Ausgangsimpulse am SEV von der Teilchenenergie abhängen.

Zur Klasse der szintillierenden Nachweisgeräte gehört noch der *Čerenkov-Zähler*. Bewegt sich ein geladenes Teilchen mit der Geschwindigkeit $v > c/n$ in einem Medium mit dem Brechungsindex n – dort also mit Überlichtgeschwindigkeit – so strahlt es Licht ab (Čerenkov-Effekt). Die Richtung des abgestrahlten Lichtes ist von der Geschwindigkeit des geladenen Teilchens abhängig.

5.4.4.2. Ionisationskammer, Zählrohr. Durchdringen geladene Teilchen Materie, so verlieren sie den größten Teil ihrer Energie durch Ionisation. Die Zahl der erzeugten Ladungsträgerpaare (Elektron + Ion) kann deshalb – wenn das Teilchen seine Energie vollständig abgibt – aus der kinetischen Energie W (Teilchen) und der mittleren Ionisationsenergie $\overline{W}$ (Ionisation) errechnet werden. Beispielsweise erzeugen Teilchen mit 1 MeV Energie in Luft

$$\frac{W\,(\text{Teilchen})}{\overline{W}\,(\text{Ionisation, Luft})} = \frac{1\,\text{MeV}}{32\,\text{eV}} \approx \begin{array}{l} 3 \cdot 10^4\,\text{Ionen} \\ \text{bzw. Elektronen.} \end{array}$$

Erreicht man, daß die entstandenen Ladungsträgerpaare nicht sofort rekombinieren, sondern in einem elektrischen Feld auseinander laufen, so sind sie zur Energiebestimmung nutzbar. In gasgefüllten Ionisationskammern sind zwei gegenüberliegende Wandungen als Kondensatorplatten ausgebildet. Wird an sie eine Spannung gelegt, bewegen sich die positiv geladenen Ionen zur einen und die Elektronen zur anderen Platte hin. An den Platten bewirkt die sich dort vollziehende Ladungskompensation einen Stromimpuls in der Spannungszuführung, der die ionisierende Strahlung charakterisiert. Ob alle erzeugten Elektronen bzw. Ionen die Platten erreichen, hängt von der angelegten Spannung ab. Erreichen sie genau alle, wird auch etwas mehr Spannung diesen Sachverhalt nicht ändern. Entsprechend der immer gleichen Anzahl von kompensierten Ladungsträgerpaaren ist der erzeugte Stromimpuls immer gleich groß – man arbeitet in „Sättigung". Bei geringerer Spannung rekombiniert ein Teil der Ladungsträger, ohne

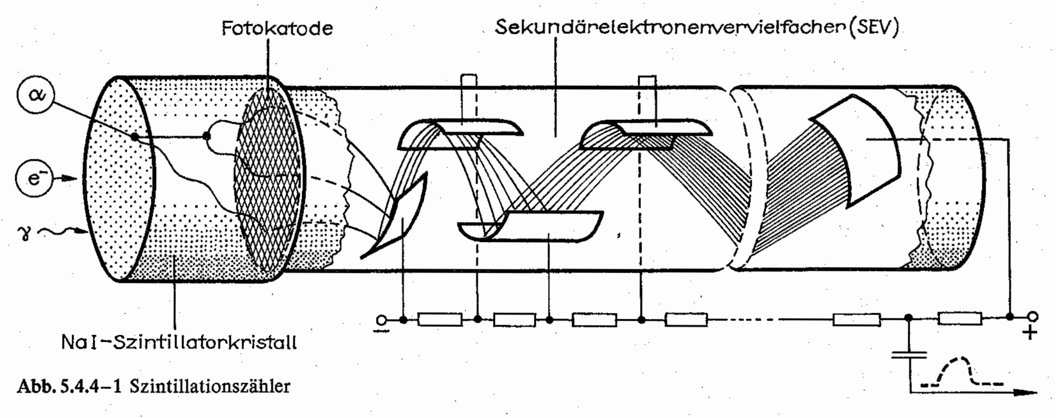


Abb. 5.4.4–1 Szintillationszähler

die Platten zu erreichen, bei höheren Spannungen werden irgendwann die Ionen im absaugenden Feld derart beschleunigt, daß sie selbst auch ionisieren und so die Anzahl der Ladungsträgerpaare ansteigt. Diese Stoßionisation ist ein erwünschter Verstärkungseffekt für den Stromimpuls, solange dieser proportional der primären Ionisation bleibt und damit eine Bestimmung der im Kammervolumen abgegebenen Energie der zu charakterisierenden Teilchen noch möglich ist. Wird die Spannung über diesen Bereich hinaus noch gesteigert, kommt es zu einem lawinenartigen Anwachsen der Zahl immer wieder ionisierender Ionen und Elektronen, bis regelrecht ein Gasplasma gezündet wird und der Durchschlag zwischen den Kondensatorplatten mit einem z. B. speziellen Gaszusatz wieder gelöscht werden muß. Derart arbeitende Ionisationskammern werden als *Zählrohre* bezeichnet (Geiger-Müller-Zählrohr). Mit Zählrohren ermittelt man nur die Zahl der eintreffenden Teilchen. Informationen über Teilchenart oder Energie sind so nicht zu erhalten. Für den Nachweis von α-Teilchen im Zählrohr, mehr noch, zur genauen Energiebestimmung in der Ionisationskammer benötigt man sehr dünne Eintrittsfenster, um den Energieverlust dort möglichst gering zu halten.

Zur Registrierung oder Messung von γ-Quanten kann man das Fenster mit einer Bleischicht für andere Teilchen undurchdringbar machen. Für Messungen an Neutronen müssen entweder – etwa in einer Paraffinschicht – Rückstoßprotonen erzeugt und diese dann vermessen werden, oder es können – etwa mit Bortrifluoridgas – über (n, α)-Reaktionen, die α-Teilchen nachgewiesen werden.

Ionisationskammern für Teilchen der Hochenergiephysik haben ein großes Volumen und sind von Ebenen paralleler Drähte durchzogen. Die Ebenen haben zueinander den gleichen Abstand, und ihre Drähte verlaufen abwechselnd um 90° versetzt. Die entlang der Bahn eines das Volumen durchdringenden Teilchens entstehenden Ladungsträger werden zu den jeweils nächstliegenden Drähten gezogen. Eine Analyse der impulsliefernden Drähte ergibt den Verlauf der Bahn des Teilchens.

5.4.4.3. Festkörperdetektoren. Man kann als aktives Medium zur Abbremsung ionisierender Teilchen statt Gas, wie im Falle der Ionisationskammer, Festkörpermaterialien,

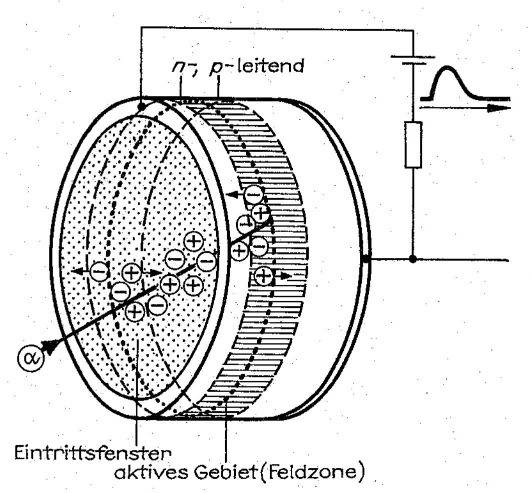


Abb. 5.4.4–2 Halbleiterdetektor für ionisierende Teilchen

wie Silicium oder Germanium, verwenden. Das hat zum einen den Vorteil, durch die um drei Größenordnungen höhere Dichte mit wesentlich kleineren Volumina auszukommen, zum anderen werden durch die geringen mittleren Ionisationsenergien von W (Ionisation) = 3,5 eV für Silicium bzw. 2,8 eV für Germanium noch um eine Größenordnung mehr Elektron-Ion-Paare bei gleicher Energieabsorption erzeugt. Das Material für Festkörperdetektoren muß für den Ladungstransport Halbleitereigenschaften aufweisen, die Funktionsweise ist aber sonst die gleiche wie bei der Ionisationskammer. Das Wirkprinzip ist in Abb. 5.4.4–2 dargestellt. Das Eintrittsfenster, gleichzeitig Elektrode, besteht meist aus einer dünnen Goldschicht. Die Größe des aktiven Volumens wird für ein gutes Signal-Rausch-Verhältnis dem Meßproblem angepaßt. Germanium-Kristalldetektoren müssen gekühlt werden, da bei der geringen Ionisationsenergie schon bei Zimmertemperatur Ladungsträgerpaare ohne äußere Einwirkungen entstehen.

5.4.4.4. Spurenbildende Detektoren. Die wesentlichen Vertreter dieser Klasse von Detektoren sind Nebelkammer, Blasenkammer und Kernspuremulsion. Sie liefern eine räumliche Spur der Bahn ionisierender Teilchen oder γ-Quanten. Diese Spuren bestehen in der Fotoplatte (Kernspuremulsion) aus Schwärzungspunkten, in der Nebelkammer aus Wassertröpfchen und in der Blasenkammer aus Gasbläschen. Die Spuren kann man bei Kernspuremulsion – die gemessen an Filmmaterial sehr dick ist – im Meßmikroskop räumlich vermessen, in der Nebel- und Blasenkammer geschieht das über Fotoaufnahmen aus verschiedenen Richtungen. Die Länge der Spur eines Teilchens ist bei bekannter Teilchenart ein Maß für seine Energie W. Ermittelt man den spezifischen Energieverlust dW/dx (vgl. 5.3.1.2.), ergibt sich eine zweite unabhängige Energiebestimmung und damit die Kenntnis der Teilchenart. Aus der Lage der Spuren im Raum ergeben sich die Reaktionswinkel. Wird der Detektor in einem Magnetfeld angeordnet, so ist aus Stärke und Richtung der Bahnkrümmung Ladungsstärke und -vorzeichen bestimmbar.

Nebelkammern arbeiten mit an Feuchtigkeit übersättigter Luft. Eine Nebelbildung beginnt bevorzugt an Kondensationskeimen. Der Vorgang ist analog der Bildung von Kondensstreifen hinter Flugzeugen in großen Höhen, wo die Nebelbildung an den Abgasteilchen einsetzt. In Nebelkammern beginnt die Kondensationskeimbildung durch Ionisation entlang der Bahn ionisierender Teilchen. Die übersättigte Atmosphäre kann in verschiedener Weise erzeugt werden. In Expansionsnebelkammern wird feuchte Luft abrupt ausgedehnt, wobei durch die Temperaturerniedrigung die relative Feuchte 100 % überschreitet und vor dem „Abregnen" kurzzeitig ein spurbereiter Zustand entsteht. Einen nicht nur kurzzeitigen, sondern andauernden spurbereiten Zustand erreicht man mit der (kontinuierlichen-) Diffusionsnebelkammer. In ihr wird ein Temperaturgefälle durch starke Unterkühlung des Bodens z. B. mit Trockeneis erreicht. Im oberen Kammerteil wird Alkohol verdunstet, sinkt ab und beginnt, in die kälteren Schichten kommend, zu kondensieren. Etwas oberhalb der selbständigen Kondensation lassen sich schöne Spuren entlang der Bahn ionisierender Teilchen beobachten, deren Tröpfchen langsam absinken, wodurch die Spuren sich verbreitern und auflösen. Für Demonstrationszwecke wird ein radioaktives Präparat an einer Schnur von oben in die sensible Luftschicht eingeführt. Nebelkammern mit Methanfüllung (CH_4) eignen sich aufgrund des hohen Wasserstoffanteiles zu einer ungefähren Darstellung der Bahn von Neutronen.

Neutronen ionisieren nicht direkt, sondern stoßen Protonen, deren Spuren sichtbar sind und entlang der Bahn des Neutrons beginnen.

Blasenkammern arbeiten mit überhitzten Flüssigkeiten im Siedeverzug. Verwendet man als Flüssigkeit Wasserstoff, vereinfacht sich die Auswertung dadurch, daß die einkommenden Teilchen nur mit dieser einen, einfachsten Atomsorte wechselwirken können. Wasserstoff siedet unter Atmosphärendruck bei 20 K. Bei 27 K (= −246 °C) und 0,5 MPa bleibt Wasserstoff noch flüssig und kommt bei Druckentlastung für ungefähr 30 ms in den spurbereiten Zustand, ehe die Flüssigkeit im gesamten Volumen siedet. Verwendet werden Kammern mit einem Fassungsvermögen in der Größenordnung von 1 m³. Andere verwendete Flüssigkeiten sind beispielsweise Xenon und Propan. Um zwischen positiv und negativ geladenen ionisierenden Teilchen unterscheiden zu können, wird ein starkes Magnetfeld angelegt. Bewegte Ladungen verschiedenen Vorzeichens laufen dann auf gekrümmten Bahnen entgegengesetzter Richtung. Blasenkammern sind durch die hohe Dichte der Flüssigkeit – relativ zur Nebelkammer – und die großen realisierbaren Volumina zur Untersuchung hochenergetischer Teilchen in der Elementarteilchenphysik geeignet. Die sehr feinen Bläschenspuren werden aus verschiedenen Ebenen fotografiert, um ihre Lage im Raum bestimmen zu können.

Funkenplattenzähler (*Streamer-Kammer*) stellen eine „Billigvariante" spurbildender Nachweisgeräte für hochenergetische Teilchen dar. Sie bestehen aus einem Paket parallel zueinander in cm-Abständen angeordneten Metallplatten in Edelgasatmosphäre. Zwischen benachbarten Platten wird kurzzeitig eine Spannung von einigen 10 kV angelegt. Durchqueren ionisierende Teilchen dieses Plattenbündel im Moment anliegender Spannung, so reichen die gebildeten Ionen aus, einen elektrischen Durchschlag zwischen den Platten hervorzurufen. Diese Funkenspur wird fotografiert und analysiert.

Kernspuremulsionen sind speziell für die verschiedenen Teilchenarten, etwa γ-Quanten, Elektronen oder α-Teilchen sensibilisierte Fotoplatten. Während normales Filmmaterial eine nur sehr dünne fotografische Schicht enthält, haben Kernspurplatten ≈ 1 mm dicke Emulsionen. In diesen bilden sich durch Ionisation längs der Bahn der durchdringenden Teilchen Schwärzungskeime, die beim Entwickeln der Platten zu einer im Mikroskop sichtbaren Spur verstärkt werden.

Kernspuremulsionen sind unterlagenfrei zu einem dicken Paket stapelbar. Die einzelnen Schichten werden erst nach der Bestrahlung auf Glasplatten aufgezogen, entwickelt und fixiert. Die in der einzelnen Schicht auch in ihrer räumlichen Lage auszuwertenden Spuren sind durch viele Schichten hindurch verfolgbar und zu vermessen.

5.4.5. Spektrometrie

Es ist Aufgabe der Spektrometrie, einen ganzen Bereich einer Meßgröße zu erfassen, z. B. die Messung der Energieverteilung der Strahlung einer Lichtquelle (das Spektrum einer Leuchte) oder der Geschwindigkeitsverteilung der Elektronen beim Zerfall eines Nuklides. Spektrometrie beinhaltet die Ermittlung eines Spektrums, einer Meßkurve und nicht nur eines einzelnen Meßpunktes. Typische Spektrometer bestehen aus einem oder mehreren Dispersionselementen und dem Detektor. Das Dispersionselement zerlegt die Strahlung nach der gewünschten Meßgröße, und der Detektor registriert die Verteilung. Zum Beispiel kann mit einem Prisma (Dispersionselement) die Intensität des Sonnenlichtes wellenlängenabhängig zerlegt werden, so daß auf einer Fotoplatte (Detektor) dann das Spektrum entsteht. Es kann auch mit einem Fotoelement das Spektrum abgefahren und registriert werden. Abb. 5.4.5−1 zeigt das Prinzip eines derartigen Spektrometers. Es ist nicht sehr effektiv, das Spektrum mit einem die einzelnen Energiewerte durchfahrenden Spalt aufzunehmen. Es wurden technische Lösungen gefunden, die Signale des gesamten Spektrums gleichzeitig zu verarbeiten.

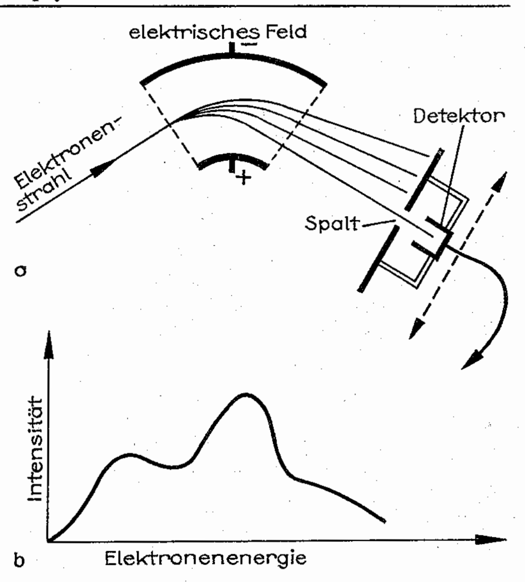


Abb. 5.4.5−1 (a) Spektrometerprinzip, (b) Spektrum (Energieverteilung)

5.4.5.1. Energieauflösende Detektoren. Detektoren können direkt eine zur Energie der zu analysierenden Teilchen proportionale Impulshöhe abgeben und dadurch die Messung einer Energieverteilung ermöglichen (vgl. 5.4.4.2.). Im Detektor entsteht eine der Teilchenenergie proportionale Anzahl von Elektron-Ion-Paaren, die zu einer energieabhängigen Impulshöhe führt. Die Impulshöhe wird in einem „Impulshöhenanalysator" digitalisiert (d. h. in eine meist vorgebbare Anzahl von Höhenstufen zerlegt), so daß z. B. in einem nachgeschalteten Vielkanalanalysator oder in den Speichern eines Rechners durch die Anzahl der Impulse in den einzelnen Höhenstufen (Kanälen) das Spektrum entsteht.

Energieauflösende Detektoren erreichen nicht das Auflösungsvermögen guter Dispersionselemente, wie etwa das des Zylinderkondensators oder eines magnetischen Sektorfeldes (vgl. 7.6.2.), sind aber durch ihre Einfachheit sehr bequem handhabbar. Sie werden z. B. als Si- oder Ge-Halbleiterdetektoren in verschiedenen Ausführungsformen verwendet, etwa als Scheiben von einigen Zentimetern Durchmessern mit einer dem Meßproblem angepaßten Dicke von einigen Millimetern. Die Dicke (genauer, die der aktiven Zone) sollte mindestens so groß wie die maximale Reichweite der nachzuweisenden Teilchen sein, um ihre Energie messen zu können. Ist die Dicke viel größer, wird dadurch

nur das Untergrundrauschen des Detektors unnötig vergrößert. Energieauflösende Halbleiterdetektoren werden vom Tage ihrer Herstellung an mit flüssigem Stickstoff gekühlt gelagert und betrieben, um die Kristallperfektion zu erhalten und damit das Auflösungsvermögen negativ beeinflussenden Prozesse möglichst zu unterbinden. Gute Halbleiterdetektoren erreichen ein Auflösungsvermögen um 100 eV. Es ist um Größenordnungen schlechter als das mit Spektrometeranordnungen erreichbare, aber für viele Aufgaben ausreichend, z.B. für die Elementanalyse mittels Zusatzeinrichtungen an Elektronenmikroskopen.

5.4.5.2. Teleskopanordnungen. Kombiniert man mehrere Halbleiterdetektoren, so sind über die einfache Registrierung und auch Energiebestimmung von Teilchen hinaus zusätzliche Informationen erhältlich. Abb. 5.4.5–2 zeigt eine Anordnung von drei Detektoren. Im ersten Detektor, der sehr dünn gehalten wird, soll nur ein Bruchteil der Energie W der Teilchen (dW/dx) absorbiert werden. Im Detektor 2 wird die gesamte Teilchenenergie aufgenommen. Alle Teilchen, die in beiden Detektoren 1 und 2 nachgewiesen werden, können nur aus dem Raumwinkel α kommen. Bei einer Schaltung der Detektoren 1 und 2 in der Art, daß nur Impulse weiterverarbeitet werden, die etwa gleichzeitig entstanden, liegt eine *Koinzidenzschaltung* vor. Spricht in solch einem Zeitintervall noch Detektor 3 an, so ist die Energie des Teilchens nicht vollständig in Detektor 2 abgegeben – nicht ordentlich erfaßt worden. Eine Schaltung der Detektoren derart, daß in diesem Falle die Impulse aus 1 und 2 nicht weiterverarbeitet werden, bezeichnet man als *Antikoinzidenzschaltung*.

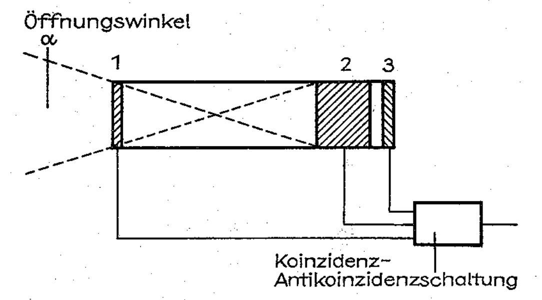


Abb. 5.4.5–2 Detektorteleskop mit drei Detektoren

Detektor 1 registriert den spezifischen Energieverlust (vgl. 5.3.1.4.), eine Meßgröße, die in Verbindung mit der Gesamtenergie eines Teilchens auch zur Bestimmung der Teilchenart dienen kann. Der spezifische Energieverlust ist eine Funktion von Energie, Masse und Ladung, womit bei bekannter Energie (von Detektor 2 registriert) in der Regel die Masse des Teilchens und damit die Teilchenart identifiziert werden kann.

Detektorteleskope werden zur Messung der Winkelverteilungen von Reaktionen um eine Meßprobe schwenkbar angeordnet.

5.4.5.3. β-Spektrometer. Elektronen sind als Bestandteile der Atome, Moleküle und des Festkörpers in einzigartiger Weise geeignet, die Stoffe charakterisierende Informationen zu liefern. Auch wegen der guten Möglichkeiten, Elektronenstrahlen erzeugen und besser noch als Licht mit Linsen beeinflussen, lenken zu können, werden sie in mannigfaltigster Form zur Analyse verwendet. Es wurde eine große Zahl von β-Spektrometertypen entwickelt. Be-

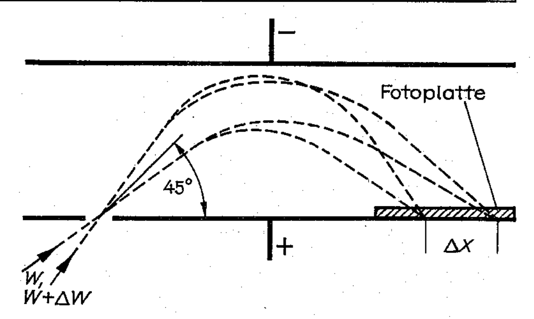


Abb. 5.4.5–3 Plattenkondensator-Spektrometer

züglich elektrostatischer Anordnungen ist der Parallelplattenkondensator die einfachste Form (Abb. 5.4.5–3). Im homogenen Feld eines Plattenkondensators wird ein unter einem Winkel von 45° eintretender divergenter Strahl von Elektronen einer Energie wieder fokussiert. Da die Elektronen verschiedener Energie W und $W + \Delta W$ nach Durchlaufen des Feldes verschiedene Endpunkte x und $x + \Delta x$ erreichen, kann die Energieverteilung z.B. auf einer Fotoplatte registriert werden. Alle Spektrometertypen, wie die mit konzentrischen Halbkugeln, mit einem Zylinderspiegel, mit verschiedenen Normallinsenanordnungen oder mit gekreuzten elektrischen und magnetischen Feldern (Wien-Filter) oder das Magnet-Prisma, erreichen in etwa die gleiche Energieauflösung von 0,1 % bzw. 1 meV.

β-Spektrometer arbeiten im Hochvakuum, um Streuungen und Reaktionen der Elektronen mit dem Restgas unwahrscheinlich werden zu lassen. Elektrostatische Spektrometer höchster Auflösung erfordern Ultrahochvakuum, um die Kontamination (Verunreinigung) an den Dispersionselementen gering zu halten, um das Feld zwischen den Kondensatorplatten nicht zu verändern. Die Spektrometer müssen sorgfältig gegen äußere, die Teilchenbahnen beeinflussende Störungen abgeschirmt werden. Das betrifft insbesondere das erdmagnetische Feld, wobei oft eine Kompensation des Feldes z. B. mit geregelten, die Apparatur umgebenden sogenannten Helmholtz-Spulen günstig ist.

β-Spektrometer werden als Zusatzeinrichtungen zu Elektronenmikroskopen, Rasterelektronenmikroskopen und als Bestandteil von Oberflächenanalysegeräten eingesetzt und liefern Beiträge zur atomaren Zusammensetzung, zu Bindungsverhältnissen und zur Struktur insbesondere von Festkörpern.

5.4.5.4. Massenspektrometer sind Massentrenner. Sie zerlegen einen Ionenstrahl in die einzelnen Massekomponenten und bestimmen so z.B. die relative Häufigkeit von Isotopen. Um eine Massenanalyse durchführen zu können, muß das Material, wenn es nicht gasförmig vorliegt, verdampft und ionisiert werden. In elektrostatischen oder/und magnetischen Feldanordnungen (vgl. 2.6.4.) erfolgt dann die masseabhängige Zerlegung. In ihnen durchlaufen die Ionen nacheinander ein elektrisches und ein magnetisches Sektorfeld. Es lassen sich Bedingungen für die Größe der Felder und die Winkel der Sektoren finden, bei denen Doppelfokussierung erreicht wird. Das bedeutet einmal, daß Ionen, die mit nicht allzu großem Divergenzwinkel in die Anlage einlaufen, in einem Punkt fokussiert werden – jeweils verschiedene Punkte für verschiedene Massen. Zum anderen werden auch Ionen einer Masse, aber unterschiedlicher Geschwindigkeit (was bei der Ionisierung nicht zu vermei-

den ist), fokussiert. Fehlt die Geschwindigkeitsfokussierung, können höchstens noch Massen, die sich um ein Tausendstel unterscheiden, getrennt werden. Das Auflösungsvermögen $M/\Delta M$ (M und $M+\Delta M$ bedeuten zwei vom Gerät gerade noch trennbare Massen) kann für doppelfokussierende Spektrometer 10^6 erreichen.
Massenspektrometer werden auch zur Isotopentrennung eingesetzt. Diese magnetischen Massentrenner haben nur eine geringe Effektivität. Sie erreichen derzeit Trennraten von einigen Milligramm je Minute.

5.4.5.5. Flugzeitmessungen sind vorteilhaft zur Bestimmung der kinetischen Energie von Neutronen durchzuführen. Vom Prinzip her ist es die ursprünglichste Art, Geschwindigkeiten zu messen. Man mißt die für einen bestimmten Weg benötigte Zeit. Das Startsignal kann dem Kollimator entnommen werden. Das ist ein schnell rotierender Zylinder mit einem Loch oder Spalt parallel zur Drehachse, mit dem der Neutronenstrahl in Impulse zerlegt wird. Mit der Messung kürzerer Zeiten erweitert sich der Meßbereich zu höheren Energien hin, und je genauer die Zeitmessung erfolgt, um so genauer ist die Energiebestimmung.

5.4.5.6. Radiografie. Radiografische Verfahren nutzen in der Regel die Eigenschaft ionisierender Teilchenstrahlen, fotografisches Material zu schwärzen. Das gegenüber dem Licht bzw. den Röntgen-Strahlen verschiedene Absorptionsvermögen ist dabei von Bedeutung. Eine dem Röntgen sehr ähnliche Methode ist die Neutronenradiografie. Statt der Röntgen-Strahlen werden Neutronen verwendet. Sie haben ein höheres Durchdringungsvermögen. Einige Elemente, wie Wasserstoff, Bor und Cadmium, absorbieren gut Neutronen. Diese Elemente und ihre Verbindungen (z. B. H_2O und Hydroxide) sind mittels der Neutronenradiografie gut nachweisbar.
In der Autoradiografie wird die aus einem Körper kommende Strahlung zur Schwärzung von Fotoplatten genutzt. Es ergeben sich Bilder der räumlichen Strahlungsverteilung. Derart kann man den Weg radioaktiver Substanzen in Gewächsen, ihre Umsetzung und den Einbau verfolgen. Es sind auch Autoradiogramme von erst im Material aktivierter Elemente – etwa mittels eines Kernreaktors – möglich. Eine sehr hohe räumliche Auflösung (Flächenverteilung) erhält man durch direkten Kontakt des Filmmaterials mit dem zu untersuchenden Körper.

5.4.6. Dosimetrie, Strahlenschutz

5.4.6.1. Dosismessung. Die Dosimetrie beschäftigt sich mit der quantitativen Erfassung von Strahlenwirkungen. Insbesondere für Strahlenschutzmaßnahmen ist es wichtig, die inkorporierte Strahlung unter Berücksichtigung ihrer Wirkung genau zu kennen. Eine Meßgröße ist die *Energiedosis* D, die Einheit ist das Gray (Gy), $1\,\text{Gy} = 1\,\text{J}\,\text{kg}^{-1}$. Es ist die Energiedosis, die der durch ionisierende Strahlung einer homogen verteilten Materienmenge von 1 kg gleichmäßig zugeführten Energie von 1 J entspricht. Nicht mehr gebräuchlich ist die Einheit rad (*r*adiation *a*bsorption *d*ose), $1\,\text{rad} = 10^{-2}\,\text{Gy}$.
Da die Strahlenwirkung für die verschiedenen Strahlenarten und auch energieabhängig bei gleicher Energiedosis verschieden ist, werden die Unterschiede durch verschiedene Qualitätsfaktoren berücksichtigt. Das Produkt aus Qualitätsfaktor Q und Energiedosis D wird *Äquivalentdosis* oder *Dosisäquivalent* H genannt: $H = QD$. Der Qualitätsfaktor wird oft noch als RBW-Faktor (relative biologische Wirksamkeit) bezeichnet. Die Einheit der Äquivalentdosis ist 1 Sievert (Sv). Die entsprechende nicht mehr gebräuchliche Einheit war das rem (*r*oentgen *e*quivalent *m*edical), $1\,\text{rem} = 10^{-2}\,\text{Sv}$.
Tab. 5.4.6–1 gibt Qualitätsfaktoren für verschiedene Strahlenarten an. Die Angaben sind nur grobe Richtwerte. Zum einen können die Energieabhängigkeiten und zum anderen die Strahlenwirkungen für verschiedene Gewebearten verschieden sein. Beispielsweise ist für eine Bestrahlung der Augenlinsen mit Protonen $Q = 30$ zu verwenden. Bezugsbasis für die Qualitätsfaktoren ist die Wirkung einer Röntgen-Strahlung von 200 keV Energie. Die zeitliche Ableitung der Äquivalentdosis H bzw. der Energiedosis D ist die Äquivalentdosisleistung P_H bzw. Energiedosisleistung P_D. Die Einheit ist $1\,\text{Sv}\,\text{s}^{-1}$ bzw. $1\,\text{Gy}\,\text{s}^{-1}$. Um eine Vergleichsbasis für die in der Strahlenschutzverordnung angegebenen Grenzwerte zu haben, sei angeführt, daß die Äquivalentdosisleistung durch die natürliche Umweltaktivität $0{,}001\,\text{Sv}\,\text{a}^{-1}$ beträgt, wozu die kosmische Strahlung $\approx 25\,\%$, die radioaktiven Stoffe der Umwelt $\approx 50\,\%$ und die im menschlichen Körper eingebauten Substanzen $\approx 25\,\%$ beitragen. Die durch zivilisatorische Einflüsse erzeugte Strahlenbelastung entspricht etwa der durch die natürliche Strahlenbelastung gegebenen Menge, 95 % entstammen der Röntgen-Diagnostik.

Tab. 5.4.6–1 Qualitätsfaktoren Q

Strahlenart	Q in $\text{Sv}\,\text{Gy}^{-1}$
Röntgen- bzw. γ-Strahlung	1,0
Elektronen	1,0
Protonen	8–10
α-Teilchen	15–20
thermische Neutronen	3
schnelle Neutronen	10

Im Umgang mit radioaktiven Substanzen ist noch der Begriff der *Aktivität* als Maß für die Menge der in der Zeiteinheit zerfallenden bzw. sich umwandelnden Kerne von Bedeutung. Die Einheit der Aktivität ist das Becquerel (Bq). 1 Bq ist die Aktivität, bei der sich im Mittel in der Zeit von 1 s 1 Atomkern eines radioaktiven Nuklids umwandelt. Nicht mehr gebräuchlich ist die Einheit Curie (Ci), der die Aktivität von 1 g Uranium $= 3{,}7 \cdot 10^{10}\,\text{Bq}$ zugrunde liegt.

Dosimeter. Zum Studium der Wirkung radioaktiver Strahlen (Strahlenwirkung) auf lebende Organismen, insbesondere den Menschen, benötigt man genaue Dosismengen. Eine hohe Meßgenauigkeit mit möglichst absoluter Sicherheit des Funktionierens der Meßgeräte ist notwendig, um die Strahlenschutzvorschriften einhalten zu können. Es wurde eine große Mannigfaltigkeit von Dosimetern entwickelt.
Eine Klasse von Dosimetern beruht auf strahleninduzierten Festkörperreaktionen. Einige Festkörper sind in der Lage, über lange Zeiträume und zuverlässig einen Teil der eingestrahlten Energie zu speichern. Sie kann zur Energiedosisbestimmung freigesetzt werden, was zumeist durch Erwärmung (thermisch stimuliert) möglich ist. Im Folgenden wird als Beispiel ein auf Thermolumineszenz basierendes Dosimeter beschrieben.
Sogenannte *Speicherphosphore* speichern Energie ionisierender Strahlung dadurch, daß einige der Elektronen, die

durch die Strahlung aus dem Valenzband in das Leitungsband gehoben werden, in Haftstellen gelangen, wo sie relativ fest gebunden sind. Erwärmt man die exponierten Speicherphosphore auf ≈ 100 bis 300 °C, kehren die in den Haftstellen lagernden Elektronen teilweise unter Lichtemission in das Valenzband zurück. Der Lichtstrom dieser Thermolumineszenz ist der aufgenommenen Energiedosis proportional. Das Material hat eine ausgezeichnete Langzeitstabilität, die Detektoren können über Monate hinweg verlustfrei akkumulierend arbeiten. Nach Entleerung der Haftstellen beim Ausheizen erreicht das Material wieder den Ausgangszustand und ist erneut als Dosimeter einsetzbar. Zur Auswertung gibt es automatisch arbeitende Geräte von der Größe eines Kofferradios, die auch netzunabhängig betrieben werden können. Typische verwendete Speicherphosphore bestehen aus Lithiumfluorid (LiF), mit Mangan aktiviertem Calciumfluorid (CaF_2) bzw. $Li_2B_4O_7$ oder aus mit Dysprosium aktiviertem Calciumsulfat ($CaSO_4$). Das Material wird als Pulver oder zu Tabletten gepreßt verwendet. Durch ihre geringen Abmessungen sind die Dosimeter bequem an den zu kontollierenden Orten unterzubringen. Selbst als Fingerringe gibt es sie. Thermolumineszenzdosimeter erfassen einen Energiedosisbereich von 10^{-6} bis 10^4 J kg^{-1}.

5.4.6.2. Grundregeln des Strahlenschutzes. Der Mensch besitzt keine Sinnesorgane, radioaktive Strahlung zu empfinden. Zur Kontrolle der Strahlenbelastung wurden Dosimeter (vgl. 5.4.6.1.) für verschiedene Anwendungsfälle entwickelt: zum direkten Ablesen, zur täglichen Energiedosismessung und für über längere Zeiträume zu addierende Strahlenbelastungen. Letztere Dosimeter werden zumeist als Plaketten an der Kleidung getragen.
Drei Grundregeln gelten für den Aufenthalt in Strahlenfeldern:

1. maximal möglichen Abstand einnehmen. (Selbstverständlich ist das Gebot, nie in direkten Kontakt mit radioaktiven Präparaten zu kommen.) Bei als punktförmig anzusehender Strahlungsquelle nimmt die Belastung mit dem Quadrat der Entfernung ab. Mitunter kommt man dadurch durch die Absorption der Luft sogar außerhalb der Reichweite der Strahlung überhaupt. Beispielsweise haben Protonen mit einer Energie von 5 MeV eine mittlere Reichweite von knapp 0,5 m in Luft.
2. Abschirmmöglichkeiten nutzen. Da beispielsweise Blei in einem großen Energiebereich Teilchenstrahlung um fast drei Größenordnungen stärker absorbiert als Luft, sind Abschirmungen leicht zu realisieren. In der Medizin werden beim Röntgen empfindliche Körperteile, die der Strahlung ausgesetzt wären und deren Abbildung nicht erforderlich ist, mit Blei abgeschirmt.
3. Aufenthaltszeiten im Strahlenbereich so gering wie möglich halten. Die aufgenommene Energiedosis ist proportional der Zeit, also um so geringer, je geringer die Aufenthaltszeit im Strahlenfeld. In Notfällen wird durch den kurzzeitigen Einsatz mehrerer Personen – wobei jeder einzelne unterhalb der höchstzulässigen Dosis bleibt – die Gefährdung einer einzelnen Person, die sich länger im kritischen Bereich aufhalten müßte, ausgeschlossen.

Derzeit existiert keine gesicherte Theorie, aus der für geringe Strahlenbelastungen die zu erwartenden Schädigungen genau vorhergesagt werden können. Experimentelle Fakten beziehen sich auf sehr starke Strahlenbelastungen. Geringe Dosen sind nicht direkt in Bezug zu Schädigungen zu bringen, da die Wirkungen zu schwach in Erscheinung treten bzw. zu selten sind, um statistisch gesicherte Aussagen zu bekommen. Bei der Extrapolation der Verhältnisse von starker Belastung zu geringen Dosen hin vermutet man, zu hohe Wirkungen vorherzusagen. Es ist anzunehmen, daß eventuell körpereigene Ausheilmechanismen existieren, die es bei sehr geringer Belastung zu keiner Schädigung kommen lassen.

5.5. Aufbau der Atome

5.5.1. Einige Eigenschaften der Atome

Es gibt in der Physik der Atome – ob es die Physik der Atomhülle, des -kernes oder der Elementarteilchen ist – Anschauungsschwierigkeiten. Bei den Elementarteilchen ist das nicht mehr so wesentlich. Wer sich bis dorthin durchstudiert hat, hat sich an vieles Ungewöhnliche gewöhnt. Die Anschauungsschwierigkeiten beim ersten Eindringen in die Physik der kleinsten Längen und Massen, kürzesten Zeiten und größten Zahlen ist im Sinne des Wortes natürlich. Wir verstehen unsere direkt erfaßbare Umwelt, die unsere Entwicklung, die ganze Menschwerdung beeinflußte, unsere Empfindungen und Gefühle herausbildete. Aber dieser Horizont ist gemessen an dem, was wir heute über unsere insgesamt existierende Umwelt wissen, sehr beschränkt. Es ist nur ein minimaler Anteil des real Existierenden, der unsere Sinne und Gefühle und somit unser erstes Verständnis herausbildete. Das Auge als wichtigstes Sinnesorgan erfaßt nur einen kleinen Teil des Spektrums der elektromagnetischen Wellen. Auch das Reich der großen Zahlen liegt weit außerhalb der eigentlichen Vorstellungswelt. Ein Beispiel soll dies verdeutlichen: Ein Glas Wasser enthält $\approx 3 \cdot 10^{24}$ Atome. Angenommen, sie wären markierbar, dann könnte man sie im Gedankenexperiment gleichmäßig in alle Wasser der Erde verteilen, gut in den Ozeanen verrühren. Der Vergleich geht auf KELVIN zurück, der weiter ausführte, daß, wenn man nach gleichmäßiger Durchmischung irgendwo ein Glas Wasser schöpft, in diesem dann immer noch ≈ 200 der markierten Atome zu finden sein würden.
Alles das besagt, daß wir nicht erwarten können, die Welt aus unseren Gefühlen heraus, von unserer engen Begriffswelt her begreifen zu können. Es ist nicht verwunderlich, in den fernen Räumen des Weltalls und den gleichermaßen fernen Atomen auf Dinge zu stoßen, die sozusagen sinnesfremd sind. Beispiel dafür ist der Welle-Teilchen-Dualismus (vgl. 5.5.4.), der es nicht zuläßt, die Größe der Atome genau anzugeben, und wer gar die Größe des Elektrons verstehen will, begibt sich leicht in den Bereich der Spekulation (vgl. 5.5.3.3.). Die Atomphysik zu betreiben, setzt eine saubere, wissenschaftlich gebildete Logik voraus und die Handhabung solch leistungsstarker Hilfsmittel wie die Quantentheorie. Die Quantentheorie führt, fundiert auf wenige, gut geprüfte Grundannahmen, zur sicheren Behandlung auch komplizierter Effekte.

5.5.1.1. Atomismus in der kinetischen Gastheorie. Die Gesetze der klassischen Physik atomistisch zu verstehen, war eine der wesentlichen Aufgaben der Physiker vergangener Generationen und ist es z. B. im Verständnis vieler Eigenschaften von Festkörpern noch heute. Erstmals gelang dies in der kinetischen Gastheorie. Die Grundgleichung der kinetischen Gastheorie $pV = 2/3\ NW_{kin}$ verbindet die Größe der Vorstellung über das Gas als ein Kontinuum, nämlich

dem im Volumen V herrschenden Druck p mit den Größen der atomistischen Vorstellung von miteinander wechselwirkenden Teilchen. Im Teilchenbild genügt zur Erklärung des Druckes die Angabe der Zahl N der Teilchen und ihrer mittleren kinetischen Energie W.

Die Vorstellung geht davon aus, daß die Moleküle nur miteinander stoßen, sich sonst aber kräftefrei bewegen. Das Gesamtvolumen aller Atome hat noch sehr klein zu sein gegen das Volumen, in dem sie sich aufhalten (ideales Gas). Die Stöße erfolgen gemäß denen harter Kugeln der klassischen Mechanik. Der Gasdruck ist somit Folge der Teilchen-Wand-Wechselwirkung, der Molekülstöße gegen die Gefäßwand. Die Ableitung dieser Grundgleichung ist in allen einführenden Physiklehrbüchern zu finden und gehört auch zum Lehrstoff der zum Abitur führenden Schulen.

Die Teilchenvorstellung ist selbst in so einfacher Handhabung sehr leistungsfähig. Als Beispiel nehme man ein Glas Erbsen und mische hinein einige Stahlkugeln und Tischtennisbälle. Simuliert man durch Schütteln die Bewegung der Teilchen, so ist das Prinzip des Archimedes über den Auftrieb von Körpern in Flüssigkeiten demonstrierbar. Die Tischtennisbälle steigen in der (Erbsen-)Flüssigkeit auf und die Stahlkugeln sinken ab. Nur ein Körper mit dem Gewicht der verdrängten Erbsenmenge schwebt.

Daß andererseits die Kontinuumstheorie brauchbar ist, liegt an der großen Zahl der Moleküle und den so oft und schnell ablaufenden Einzelreaktionen. 1 cm^3 eines Gases enthält bei 0 °C $\approx 2 \cdot 10^{19}$ Moleküle, die sich mit einer Geschwindigkeit von ≈ 1700 m/s bewegen. Jedes Molekül erleidet $\approx 10^{10}$ Stöße je Sekunde.

5.5.1.2. Die Größe der Atome. Mit einfachen Mitteln gelangt man nicht sicher zu Aussagen über die Größe der Atome, aber „kleiner als"-Aussagen sind möglich. Beispielsweise läßt sich mit dem Volumen eines Öltröpfchens, unter der Annahme, daß es sich auf dem Wasser zu einer einatomaren (mono-atomaren) Schicht entspannt, wenn man die Flächenausdehnung mißt, die Dicke der Schicht und damit der Atomdurchmesser bestimmen.

Eine sichere Methode ist die der Röntgen-Beugung. Von einem Festkörper mit regelmäßig angeordneten Atomen (Kristallgitter) lassen sich mit Röntgen-Strahlen Beugungsbilder erzeugen, die es gestatten, den Abstand der Atome genau zu bestimmen. Viele unabhängige Verfahren sind erprobt worden, die Atomgröße festzustellen. Jedes der Verfahren liefert für eine Atomsorte einen etwas verschiedenen Radius, je nach dem, welche Annahmen dem Verfahren zugrunde liegen. Die Abweichungen sind aber nicht sehr groß. Alle bestätigen die Aussage, daß die Atomradien in der Größenordnung von 10^{-10} m liegen.

Die Größe der Atome ist durch ihre Elektronenhülle gegeben. Die Ergebnisse der Wellenmechanik, etwa am Beispiel der Unbestimmtheitsrelation (vgl. 5.5.4.3.), gestatten es nicht, die Bahnradien der Hüllelektronen genau anzugeben. Aber auch in einem Festkörper ist je nach Bindungstyp der Atome ein verschiedener Radius anzugeben. Die Abhängigkeit der Atomgröße von der Ordnungszahl ist gering. Wesentlicher ist die Stellung im Periodensystem der Elemente. Relativ große Radien haben die Elemente der ersten Hauptgruppe Lithium, Natrium, Kalium, Rubidium, Caesium. Sie sind etwa doppelt so groß wie die der sechsten Hauptgruppe Sauerstoff, Schwefel, Selen, Tellur. Beispiele einiger Atomradien in 10^{-10} m: Lithium 1,34, Beryllium 0,90, Kohlenstoff 0,77, Stickstoff 0,75, Magnesium 1,30, Phosphor 1,06, Chlor 0,99.

5.5.1.3. Masse der Atome. Aus der Chemie stammt der Begriff der *relativen Atommassen*. Anfang des 19. Jahrhunderts wurden von DALTON die Massen der damals bekannten Atome in Einheiten des leichtesten der Elemente – bezogen auf den Wasserstoff mit der relativen Atommasse 1 – angegeben. Heute gibt man die relativen Atommassen bezogen auf das mit genau 12 festgelegte Kohlenstoffisotop C-12 an. Entsprechend ergibt sich die relative Molekülmasse aus der Summe der zugehörigen relativen Atommassen.

Die relativen Atommassen sind im Periodensystem der Elemente angegeben. Bei älteren Tabellen kann es sein, daß die Massen noch auf das Sauerstoffisotop O-16 bezogen sind. Um die absoluten Atommassen angeben zu können, benötigt man die Zahl der Atome in einer vorgegebenen Masse. LOSCHMIDT errechnete aus thermodynamischen Daten schon 1865 die Moleküldichte für Gase. Er erkannte, daß die Zahl der Gasmoleküle bei gleichem Druck und gleicher Temperatur konstant ist, unabhängig von der Art der Moleküle. Für ideale Gase beträgt sie bei 0 °C und 100 kPa $N_A = 6{,}022\,0 \cdot 10^{23}\ \text{mol}^{-1}$ (*Avogadro-Konstante*). Mol bezeichnet eine Stoffmenge, kurz gesagt, die der relativen Molekülmasse in Gramm (molare Masse). Umgekehrt formuliert man als SI-Basiseinheit: Das Mol ist die Stoffmenge eines Systems, das aus so vielen gleichartigen elementaren Teilchen besteht, wie Atome in 0,012 kg des Kohlenstoffisotops C-12 enthalten sind. 1 Mol füllt unter den genannten Normalbedingungen ein Volumen von 0,022 4 m^3. Daraus errechnet sich die auch gebräuchliche Loschmidtsche Zahl, die Zahl der Moleküle je Volumen: $N_L = 27 \cdot 10^{18}\ \text{cm}^{-3}$. Molare Masse dividiert durch Avogadro-Konstante ergibt die absolute Atommasse, $m_a = M_x/N_A$ (M_x = Molare Masse des Elementes x), z. B. Kupfer: $m_a = 63{,}546 \cdot 10^{-3}\ \text{kg}/6{,}022\,0 \cdot 10^{23} = 10{,}55 \cdot 10^{26}$ kg.

5.5.2. Kern und Hülle

5.5.2.1. Lenards Streuversuche. Heute ist bekannt, daß das Atom aus den Nukleonen des Kerns und den Elektronen der Hülle besteht. Der Kern ist sehr klein gegen das Atom, aber seine Nukleonen haben die überwiegende Masse. Der Kern ist positiv geladen, und die Hüllelektronen erwirken mit ihrer negativen Ladung die Neutralität des Atoms. Es war in den Anfängen der Atomphysik nicht einfach, den Aufbau der Atome zu ergründen.

Die Ionisierbarkeit der Atome, die Freisetzung von Elektronen, ließ vermuten, daß Elektronen Bausteine der Atome sind.

Eine sehr direkte und noch heute gute Methode, elementare Teilchen abzutasten, ihre Struktur zu ermitteln, ist der Stoß mit ebenfalls elementaren Teilchen. Die zu untersuchenden Teilchen können in Form einer dünnen Folie oder als Gas vorliegen. Die abtastenden Teilchen, Elektronen oder etwa α-Teilchen, durchdringen die Folie bzw. das Gas. Eine Analyse der Streuverteilung der Teilchen gestattet, bei bekannten Gesetzen der Wechselwirkung, die Größe der gestoßenen Teilchen zu ermitteln.

LENARD durchstrahlte 1903 dünne Aluminiumfolien ($\approx 3\ \mu$m dick) mit Elektronen. Er betrachtete die Anzahl der Elektronen N, die die Folie durchdrangen, relativ zur Zahl der eingeschlossenen Elektronen N_0, $N/N_0 = e^{-n\delta d}$. N nimmt mit wachsender Schichtdicke d exponentiell ab und ist im gleichen Maße abhängig von der Dichte der Streuzentren n (Atome) und ihrem Wirkungsquerschnitt δ (Trefferfläche). Das Ergebnis der Versuche ergab bei nicht

zu hohen Elektronenenergien einen Atomradius wie erwartet von $R_A \sim 10^{-10}$ m.
Bei hohen Elektronenenergien allerdings durchdrangen sehr viele Elektronen die einige tausend Atome dicke Folie. Der dabei ermittelte Wirkungsquerschnitt ergab, daß nur ein sehr kleiner Teil des Atoms noch undurchdringbar ist. Der Radius dieses Kernes betrug $R_K \sim 10^{-15}$ m. Damit war der Aufbau der Atome aus einem kleinen Kern und einer leicht durchdringbaren Hülle experimentell ermittelt.

5.5.2.2. Rutherford-Streuung. Berühmt wurden die genauen Untersuchungen mit α-Teilchen im Labor von Rutherford. Geiger und Marsden fanden dort die Kernstreuung von α-Teilchen, Rutherford entwickelte die Streutheorie. Spuren von α-Teilchen in Wilsons Nebelkammer machten den Streuvorgang 1912 sogar sichtbar. In Abb. 5.5.2–1 ist die Streuung schematisch dargestellt und die sie beschreibende *Rutherfordsche Streuformel* angegeben. α-Teilchen der kinetischen Energie W_{kin} durchdringen eine dünne Folie der Dicke d. Einzig wesentliche Wechselwirkung der relativ zu Elektronen sehr schweren α-Teilchen ist die mit den punktförmig angenommenen positiv geladenen Kernen der Atome. Das Coulombsche Gesetz ist vorausgesetzt. Die Zahl der um den Winkel ϑ in das Raumwinkelelement dΩ gestreuten Teilchen dN relativ zur Zahl der eingeschossenen Teilchen N wurde experimentell genau bestimmt und bestätigte diese Streuformel. Damit war bewiesen, daß die Atome relativ luftige Gebilde mit einem sehr kleinen Streuzentrum der Ladung Z sind. Das Streuzentrum, der Kern, mußte von einer Masse größer als die der α-Teilchen sein (Nachweis von Streuwinkeln über 90°).

Das von Rutherford in Auswertung der Experimente zu seiner Streuformel postulierte Atommodell lehnte sich an das der die Sonne umkreisenden Planeten an (Planetenmodell). Z Elektronen umkreisen in großem Abstand den fast die gesamte Masse enthaltenden, aber sehr kleinen Kern der Ladung $+Ze$. Coulombsche Anziehungskraft und Fliehkraft sind entgegengerichtet und gleich. Im bildhaften Vergleich der Größenverhältnisse gehört zu einem stecknadelkopfgroßen Kern ein Atom von 100 m Durchmesser, d.h. in dieser Entfernung kreisen noch Elektronen. Verblüffend ist in dieser Vorstellung die Härte der Festkörper. Ein Festkörper hat dicht an dicht liegende Kugelhüllen der Atome. Andererseits, könnte man die Hüllen eindrücken, würde die Erde auf einen Durchmesser von 100 m schrumpfen.

5.5.3. Eigenschaften des Elektrons

5.5.3.1. Die Elementarladung. Elektronen sind die kleinsten und leichtesten Träger der Elementarladung. Unter Elementarladung versteht man die kleinste bisher nachgewiesene Ladungsmenge. Alle größeren Ladungsmengen sind sehr genau Vielfache dieser Elementarladung. (In der Elementarteilchenphysik hat man 1964 eine Gruppe von neuen Teilchen, die Quarks, eingeführt, denen 1/3 bzw. 2/3 der Elementarladung zuzuordnen sind (vgl. 5.1.2.3.). Ihr experimenteller Nachweis steht noch aus. Freie Elektronen existieren nicht ohne weiteres, Ladungen entstehen immer paarweise und haben die Eigenschaft, sich zu kompensieren. Elektronen existieren als Bestandteil der in der Regel

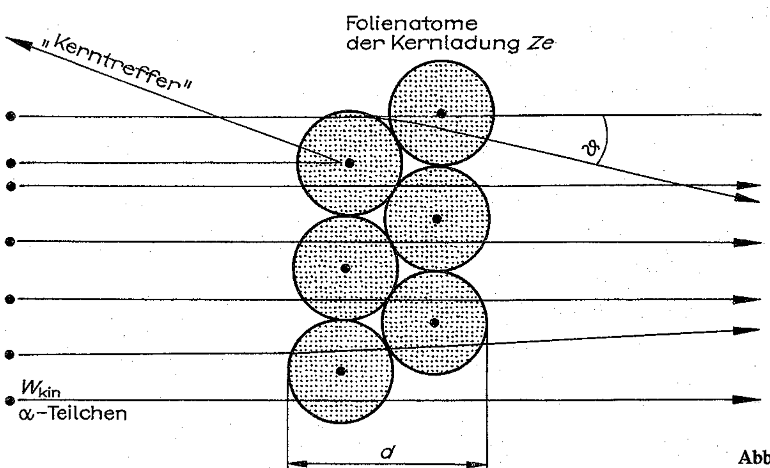


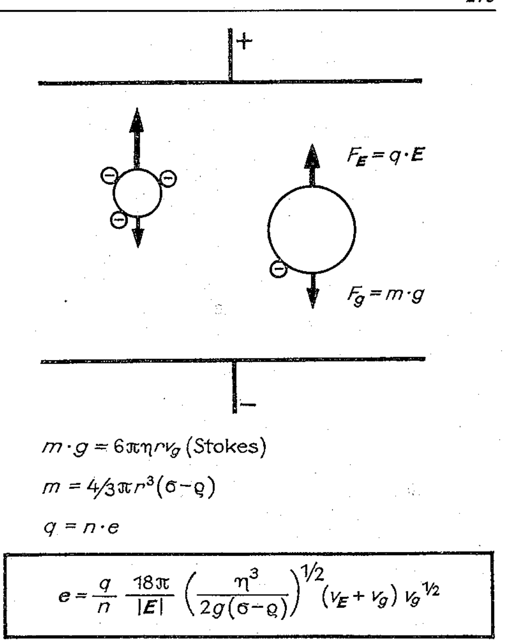
Abb. 5.5.2–1 Rutherford-Streuung

neutralen Atome. Freie Elektronen sind in mannigfaltiger Weise erzeugbar, z. B. durch Erhitzen eines Drahtes (thermische Emission von Elektronen), beim Absprühen unter hohen Feldstärken (Feldemission), durch Lichteinstrahlung (fotoelektrischer Effekt) oder durch Stoßionisation beim Zusammenprall elementarer Teilchen. Unter Ionisation versteht man die Ablösung eines oder mehrerer Elektronen vom Atom oder Molekül, das dann als positives Ion zurückbleibt.
Der Name Elektron entstammt dem Griechischen und war gleichbedeutend mit Bernstein. Bernstein hat die Eigenschaft, nach Vorbehandlung durch Reiben, leichte Gegen-

Sehr selten, aber doch sicher nachweisbar waren „Kerntreffer". Das waren Streuungen in sehr große Winkel, die nach der Streuformel unwahrscheinlich sind. Diese ersten Hinweise auf andere als Coulombsche Wechselwirkungskräfte zeigten die Wirkung der Kernkräfte an.
Während die ersten Messungen zur Rutherford-Streuung noch nicht von ausreichender Genauigkeit waren, den Wert der Kernladung $Z \cdot e$ zu ermitteln, gelang es danach dem späteren Entdecker des Neutrons, J. Chadwick, durch genaue Messung von dN und N die Kernladung zu bestimmen und als gleich der Ordnungszahl im Periodensystem der Elemente zu erkennen.

stände anzuziehen. Die Festlegung, daß das Elektron negativ geladen ist und die es elektrisch kompensierenden Teilchen positiv geladen sind, ist willkürlich. Die Kraft der elektrischen Anziehung zwischen positiv und negativ geladenen Teilchen bzw. der Abstoßung zwischen zwei Ladungsträgern gleicher Polarität wurde von de COULOMB durch genaues Experimentieren bestimmt und ist mit dem nach ihm benannten *Coulombschen Gesetz* berechenbar, $F = q_1 \cdot q_2/(4\pi\varepsilon_0 r^2)$, $\varepsilon_0 = 8{,}85 \cdot 10^{-12}\,\mathrm{AsV^{-1}m^{-1}}$.
Die Kraft F ist proportional dem Produkt der beiden Ladungsmengen q_1 bzw. q_2 und umgekehrt proportional dem Quadrat des Abstandes der Ladungsträger zueinander. Das Coulombsche Gesetz ähnelt sehr dem der Massenanziehung. Die Gravitationskraft enthält statt der Ladungen die sich allerdings nur anziehenden Massen. Die Coulomb-Kraft ist weitaus stärker als die Massenanziehung. So überwiegt die elektrostatische Abstoßung zweier Elektronen die der Massenanziehung um den Faktor 10^{41}.
Schon im ersten Drittel des vergangenen Jahrhunderts konnte man aus den Faradayschen Gesetzen der Elektrolyse auf die Größe der Elementarladung schließen. Man kannte die zum Transport einer Stoffmenge durch den Elektrolyten notwendige Elektrizitätsmenge und schloß über die Zahl der Atome der Stoffmenge auf die Größe der Ladung eines Ions. Die Faradayschen Gesetze besagen, daß zum einen die in einem Elektrolyten an der Elektrode abgeschiedene Stoffmenge den transportierten Ladungsmengen proportional ist und zum anderen, daß zum Transport von $1/n$ Mol eines n-wertigen Stoffes die Ladungsmenge von $96\,493\,\mathrm{C} = F$ (Faraday-Konstante) benötigt wird. Die Zahl der Atome der Stoffmenge 1 Mol ist durch die Avogadro-Konstante gegeben $N_A = 6{,}022\,0 \cdot 10^{23}\,\mathrm{Mol^{-1}}$. Somit ist einem einwertigen Ion die Ladungsmenge e zuzuordnen, $e = F \cdot N_A^{-1}$, $e = 96\,487\,\mathrm{C}/6{,}022\,0 \cdot 10^{23}\,\mathrm{Mol} = 1{,}602\,2 \cdot 10^{-19}\,\mathrm{C}$. C ist das Zeichen für die Ladungseinheit Coulomb (1 C = 1 As). MILLIKAN bekam für seinen sehr eleganten Nachweis der Existenz einer Elementarladung 1923 den Nobelpreis für Physik. Das Prinzip der Methode ist in Abb. 5.5.3–1 dargestellt. Ein Ion fällt im Schwerefeld der Erde abwärts. Läßt man in einer Kondensatoranordnung ein elektrisches Feld in entgegengesetzter Richtung wirken, wird je nach Stärke des elektrischen Feldes das Ion abfallen oder aufsteigen. Im Schwebefall, im Gleichgewicht der elektrischen und der Gravitationskraft, läßt sich die Ionenladung q bestimmen. Da ein Ion nicht beobachtbar ist, verwendet man möglichst einfach geladene kleine Öltröpfchen (Radius r, Masse m), die mit einem Zerstäuber gut zu erzeugen und über eine Fernrohrvergrößerung gut beobachtbar sind. Das gegen ein Ion sehr schwere Öltröpfchen hat noch den Vorteil, einer stärkeren Gravitationskraft zu unterliegen. Da der Schwebefall nicht problemlos einstellbar ist – die Brownsche Bewegung und anderes wirken sich störend aus – mißt man das langsame Fallen und Steigen der Tröpfchen. Sie fallen oder steigen mit gleichförmiger Geschwindigkeit v_g bzw. v_e (unter Gültigkeit des Stokesschen Gesetzes), wobei die größte Ungenauigkeit der Koeffizient der inneren Reibung η für Luft darstellt. Ferner ist ein Auftrieb der Tröpfchen in der Luft zu berücksichtigen (Dichte in den Tröpfchen δ, der Luft ϱ), und wenn man sehr genau experimentiert, macht sich noch eine Abweichung vom Stokesschen Gesetz bemerkbar. Der Radius der Tröpfchen ist nicht groß gegen die Weglänge zwischen den Stößen mit Luftmolekülen; ein in Abb. 5.5.3–1 nicht angeführter Korrekturterm wird notwendig.

5.5.3.2. Masse des Elektrons. Das Elektron ist das leichteste der ruhmassebehafteten elementaren Teilchen. In neuester Zeit wird auch dem Neutrino eine Ruhmasse zugeordnet. (Das Neutrino ist ein relativ seltenes, zumindest aber selten in Erscheinung tretendes Elementarteilchen, welches bei bestimmten Kernreaktionen entsteht, vgl. 5.1.2.). Der Begriff der Ruhmasse entstammt der Relativitätstheorie. Ruhmasse ist die Masse eines sich nicht in Bewegung befindlichen Teilchens oder Körpers. Die Masse vergrößert



$$m \cdot g = 6\pi\eta r v_g \text{ (Stokes)}$$
$$m = 4/3\pi r^3(\sigma - \varrho)$$
$$q = n \cdot e$$
$$e = \frac{q}{n} \frac{18\pi}{|E|} \left(\frac{\eta^3}{2g(\sigma-\varrho)}\right)^{1/2} (v_E + v_g)\, v_g^{1/2}$$

Abb. 5.5.3–1 Bestimmung der Elementarladung e nach MILLIKAN. Kleine geladene Öltröpfchen fallen im Schwerefeld der Erde und steigen im elektrischen Feld eines Kondensators

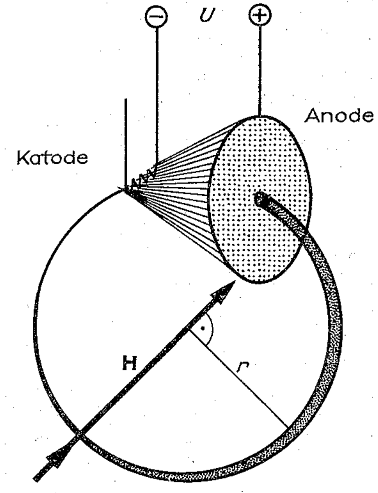


Abb. 5.5.3–2 Versuchsanordnung zur Bestimmung der Masse des Elektrons aus der spezifischen Ladung

sich mit zunehmender Geschwindigkeit, ein Effekt, der allerdings erst bei Geschwindigkeiten, die nicht mehr klein gegen die der Lichtgeschwindigkeit sind, merklich wird. Die Ruhmasse des Elektrons beträgt $9{,}1 \cdot 10^{-31}$ kg, das ist nur der 1/1836 Teil der Masse des Wasserstoffatoms.

Zur Bestimmung der Elektronenmasse kann man die Gesetze der Bewegung geladener Teilchen in elektrischen und magnetischen Feldern anwenden. Abb. 5.5.3–2 skizziert eine Versuchsanordnung zur Bestimmung der Masse des Elektrons. Hierbei mißt man die spezifische Ladung e/m, woraus sich bei bekannter Ladung e die Masse m ergibt.

In einem Vakuumgefäß mit etwas Quecksilberdampf zur Sichtbarmachung des Elektronenstrahles werden Elektronen aus einer Glühkatode K zur Anode A hin beschleunigt. Ein feiner Strahl dringt durch das Loch in der Anode. Ein senkrecht zu seiner Ausbreitungsrichtung angeordnetes Magnetfeld $\boldsymbol{H}$ zwingt die Elektronen auf eine Kreisbahn. Hat man die Beschleunigungsspannung, den Radius r der Kreisbahn und die Stärke des Magnetfeldes, so ist e/m und damit die Elektronenmasse errechenbar.

5.5.3.3. Klassischer Elektronenradius. Informationen über Größe und Gestalt erhält man allgemein aus Wechselwirkungen. Zur Bestimmung des Durchmessers beispielsweise eines Tischtennisballes mittels Schiebelehre ist der Einsatz der Kraft auf die Meßschneiden die entscheidende Information. Beim Ausmessen eines Wattebausches ist es nicht so einfach, die beginnende Wechselwirkung festzustellen. Unmöglich wird es beim Elektron. Die Kraftwirkung ist umgekehrt proportional dem Abstand zu einem wechselwirkenden Teilchen. Sie nimmt mit abnehmender Entfernung gleichmäßig zu und gibt nirgends einen Anlaß, einen Radius festzulegen. Es gibt keine Hinweise auf eine Struktur der Elektronen. Auch über den Ursprung der Ladung ist nichts bekannt. Die Einsteinsche Masse-Energie-Äquivalenz eröffnet eine Möglichkeit einer Vorstellung zur Größe. Angenommen, die gesamte Masse m des Elektrons wäre nur reine elektrostatische Energie $W_{\text{pot}}^{m} = mc^2$ (c = Lichtgeschwindigkeit), dann gibt es einen Radius r_{kl} im Coulombschen Gesetz der Wechselwirkung von Ladungen, bei dem – nur die Elementarladung mit sich selbst wechselwirkend eingesetzt – gerade W_{pot}^{m} erreicht ist, $W_{\text{pot}}^{m} = e^2/(4\pi\varepsilon_0 r_{\text{kl}})$. Zerlegt man in einem Gedankenexperiment die Elementarladung e und verteilt sie gleichmäßig auf eine Kugeloberfläche, so wird bei r_{kl} die Abstoßungsenergie der Elementarladungsteile gleich der aus der Masse des Elektrons entnehmbaren Energie mc^2. Das ist die Definition für den *klassischen Elektronenradius*. Man kann die Ladung auch homogen in einer Kugel verteilen, das ändert den Wert nicht sehr, er beträgt $r_{\text{kl}} = 2{,}8 \cdot 10^{-15}$ m (klassischer Elektronenradius). Voraussetzung ist die Gültigkeit des Coulombschen Gesetzes auch für so extrem kleine Distanzen. Die Quantentheorie lehrt allgemein, daß die Vorstellungen der klassischen Physik in diesem Größenbereich versagen können.

5.5.4. Welle-Teilchen-Dualismus

Die Jahrhunderte währende Suche nach der Natur des Lichtes konnte erst mit Aufkommen der Quantentheorie beendet werden. Nur mit ihr lassen sich alle Eigenschaften befriedigend erklären. Dabei geht es nicht nur um spezielle, nur dem Fachmann bekannte Effekte, auch die Funktion eines Lichtmessers im Fotoapparat, das Leuchten einer Glühlampe und die Existenz stabiler Atome werden erst mit der Quantentheorie verständlich. Eines ihrer Grundaxiome ist der *Welle-Teilchen-Dualismus* (Welle-Teilchen-Doppelcharakter).

Die klassische Physik unterscheidet zwischen Wellen und Teilchen. Die Wellenlehre beschreibt die einen Fälle, die Mechanik die anderen. Beides waren unvereinbare Dinge. Erst die Wellenmechanik der Quantentheorie vereinte die Wellen – und die Teilcheneigenschaften zu Teilaspekten der real existierenden Mikroobjekte – nun als *Quanten* bezeichnet. Lichtwellen bestehen aus einzelnen Lichtquanten, die auch masse- und impulsbehaftet sind. Elektronen zeigen wie alle elementaren Teilchen – heutzutage leicht nachzuweisende – Welleneigenschaften wie Beugung und Interferenz.

Im allgemeinen Sprachgebrauch hat man weiterhin den Begriff Welle, wenn nicht gerade ein einzelnes Quant gemeint ist. Man spricht auch weiterhin von Teilchen, selbst wenn, wie beim Neutrino, die Ruhmasse verschwindend klein ist. Unter Teilchen versteht man elementare Mikroobjekte mit überwiegend durch die Mechanik beschreibbaren Eigenschaften, und unter Wellen Mikroobjekte, die, wenn sie in großer Zahl auftreten, in der klassischen Physik durch die Wellenlehre beschrieben werden. Wenn beide Eigenschaften eine Rolle spielen, löst nur die Wellenmechanik das Problem der theoretischen Beschreibung.

Der in der makroskopischen Umwelt gebildete Mensch hat erst einmal Schwierigkeiten, die in der submikroskopischen Welt herrschenden physikalischen Gesetzmäßigkeiten zu begreifen. Letztendlich begründen aber die Gesetze der Mikrophysik – der Atome und ihrer Verbindungen – die der Makrophysik, der von uns zu bewältigenden Umwelt.

Grundgleichung des Welle-Teilchen-Dualismus ist die *De Broglie-Beziehung* $\lambda = h/p$ (λ = Wellenlänge, p = Impuls, h = Plancksches Wirkungsquantum). Sie ordnet den impulsbehafteten Teilchen eine Wellenlänge zu, beschreibt die Materiewellen. Gleichermaßen aber im ursprünglichen Sinne der Gleichung wird jeder Welle, beschrieben durch die Wellenlänge λ ein Impuls p bzw. eine Masse $mc = p$ (c = Lichtgeschwindigkeit) zugeordnet. (Nach Einstein gilt für die Energie einer Strahlung $W = h\nu = hc/\lambda$, und mit $W = mc^2$ ergibt sich $\lambda = h/mc$). Die die beiden Größen λ und p verbindende Konstante, das Plancksche Wirkungsquantum $h = 6{,}6 \cdot 10^{-34}$ Js läßt durch seine Kleinheit verständlich werden, warum im Alltag z. B. beim Fußballspiel die Wellenlänge des Balles keine Rolle spielt oder die Masse der Sonnenstrahlen auf uns keinen merklichen Druck ausübt. Andererseits erzeugt der Lichtdruck der Sonne den deshalb auch immer von der Sonne weggerichteten Kometenschweif und gehören Beugungsbilder mit dem Elektronenmikroskop zum selbstverständlichen Alltag vieler Wissenschaftler.

5.5.4.1. Teilcheneigenschaften elektromagnetischer Wellen. Besondere Schwierigkeiten, die Natur des Lichtes zu verstehen, brachte der äußere lichtelektrische Effekt (*Fotoeffekt*) mit sich. Erste Experimente dazu ergaben, daß die Ablösung von Elektronen aus Katodenmaterialien durch Lichteinstrahlung gemäß dem damaligen Wissensstand unverständlich war. Nach der Wellentheorie müßte die für die Ablösearbeit aufzubringende Energie nur eine Frage der Intensität des Lichtes sein. Das Experiment zeigte aber, daß unabhängig von der Intensität erst ab einer bestimmten Frequenz des Lichtes Elektronen emittiert werden, und die Energie der abgelösten Elektronen war auch von der Frequenz ν des Lichtes abhängig, $W_{\text{kin}}^{\text{Elektron}} = h\nu - W_{\text{A}}$ (Fotoeffekt, h = Plancksches Wirkungsquantum).

Offensichtlich ist die Energie des Lichtes durch $W^{\text{Licht}} = h\nu$ gegeben, und die kinetische Energie der Elektronen $W_{\text{kin}}^{\text{Elektron}}$ ist gleich der von der Katode absorbierten Energie des Lichtes, verringert um die zum Verlassen der Katode notwendige Ablösearbeit W_A.
Fünf Jahre nach diesen Experimenten fand EINSTEIN 1905 die Erklärung mit seinem Photonenmodell des Lichtes. Demnach besteht Licht aus einzelnen *Photonen* (Quanten) der Energie $h\nu$. Die Intensität des Lichtes ist durch die Zahl dieser Photonen je Zeiteinheit gegeben. Wird ein Photon absorbiert, so kann ein Elektron seine Energie übernehmen. Reicht die Energie aus, die Barriere zum Verlassen des Festkörpers zu überwinden, so können die Elektronen in einem Vakuumfotoelement als Stromfluß zur Anode nachgewiesen werden (äußerer lichtelektrischer Effekt). Die Stärke des Stromes, d. h. die Zahl der Elektronen, ist nur durch die Zahl der absorbierten Photonen gegeben. Für ein anschauliches Bild ist es oft legitim, sich das Photon als einen Wellenzug von begrenzter Länge (≈ 1 m für Licht) vorzustellen. Eine genaue Vorstellung, ein immer gültiges anschauliches Bild scheitert an unseren Sinneswahrnehmungen, denen es nicht gegeben ist, den Doppelcharakter der Quanten, Welle und Teilchen zugleich zu sein, widerspruchsfrei zu verarbeiten. Der Teilchencharakter der elektromagnetischen Wellen kommt eindrucksvoll beim *Compton-Effekt* zu Tage. Wenn dem Lichtquant eine Masse m gemäß $W = h\nu = mc^2$, also $m = h\nu c^{-2}$ zuzuschreiben ist, so müßte es möglich sein, den Stoß eines Lichtquantes mit einem Teilchen, etwa einem Elektron, auch gemäß den Gesetzen der Mechanik, mit denen z. B. der Stoß zweier Billardkugeln gerechnet wird, beschreiben zu können. COMPTON bewies dies in einem Experiment, bei dem Röntgen-Strahlen mit Atomelektronen wechselwirkten. Formuliert man in diesem Falle Energie- und Impulssatz, so ergibt sich experimentell eine exzellente Übereinstimmung mit den Stoßgesetzen, d. h. der Erfüllung von Energie- und Impulssatz. Nach dem Stoß hat sich die Energie der Wellen von $h\nu_0$ durch Frequenzänderung auf $h\nu$ verringert, und die Energiedifferenz entspricht der kinetischen Energie der vor dem Stoß als ruhend angenommenen Elektronen, $h\nu_0 = h\nu + W_{\text{kin}}^{\text{Elektron}}$; $W_{\text{kin}}^{\text{Elektron}} = mv^2/2$.
Die von Lichtquanten getroffenen Elektronen erhalten einen Rückstoß, als wären sie von Teilchen mit dem Impuls $h\nu/c$ gestoßen worden.
Die Masse eines einzelnen Lichtquantes (Photons) im Bereich des sichtbaren Lichtes ist von der Größenordnung 10^{-36} kg. Die Lichtquantenmasse nimmt in einer Weise eine Sonderstellung ein. Es handelt sich um die bewegte, die relativistische Masse. Die Ruhmasse der Quanten elektromagnetischer Strahlung muß Null sein. Da aber die Existenz dieser Quanten an die Bewegung mit Lichtgeschwindigkeit gebunden ist, ist ihre Ruhmasse nur eine Größe der Theorie.
Die Masse der Lichtquanten wurde schon zu Anfang des Jahrhunderts aus dem Lichtdruck experimentell bestimmt.

5.5.4.2. Welleneigenschaften von Teilchen. Die Vorhersage von de BROGLIE, daß Teilchen jedweder Art der Masse m, die sich mit der Geschwindigkeit v bewegen, eine Wellenlänge $\lambda_{\text{Teilchen}} = h/mv$ (*Materiewellen*, $h = 6{,}6 \cdot 10^{-34}$ Js) zuzuordnen ist, ist experimentell ausreichend bestätigt. Am einfachsten ist die De-Broglie-Relation für Elektronen nachzuprüfen. Die Kleinheit des Planckschen Wirkungsquantums h erfordert auch einen kleinen Nenner, also kleine Massen, um einfach auf ausreichend große Wellenlängen zu kommen. Allerdings ist für Elektronen schon ab geringer Energie die relativistische Massenzunahme zu berücksichtigen. Die wirksame Masse m ergibt sich aus der Ruhmasse m und der Elektronengeschwindigkeit v zu $m = m_0 \quad (1 - v^2/c^2)^{-1/2}$ (c = Lichtgeschwindigkeit, vgl. 3.1.1.3.). Beispielsweise sind Elektronen der Energie von 1 eV und 100 keV Wellenlängen von $\lambda = 2{,}4 \cdot 10^{-9}$ m bzw. $3{,}6 \cdot 10^{-12}$ m zugeordnet. Die Energie geladener Teilchen wird oft in *Elektronenvolt* (eV) angegeben. Es ist die Energie, die ein einfach geladenes Teilchen beim Durchlaufen der Spannung U durch Beschleunigung seiner Bewegung dem Feld entzieht und als Energie mit sich führt, $W_{\text{kin}} = mv^2/2 = eU$. Als Umrechnung gilt $1\text{ eV} = 1{,}602 \cdot 10^{-19}$ J. Die Wellenlängen der Elektronen mit Energien in der Größenordnung von 10 bis 100 eV sind vergleichbar mit den Atomabständen (Gitterkonstanten) in Festkörpern, so daß sich damit leicht Beugungseffekte nachweisen lassen. Einige Methoden der Festkörperanalytik bauen darauf auf. Interferenzbilder aus Elektronen- oder beispielsweise Neutronenstreuexperimenten sind identisch mit denen von Röntgenstrahlen gleicher Wellenlänge.

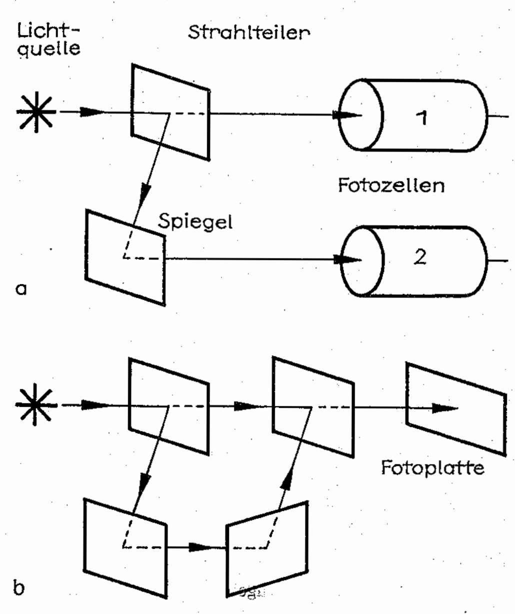


Abb. 5.5.4–1 Diracs Spiegelversuch, (a) Beweis der Lokalisierbarkeit und (b) Beweis der Interferenz eines Photons

5.5.4.3. Dirac'scher Spiegelversuch. Experimente zur Natur der elementaren Quanten offenbaren noch heute, wie schwer gewohnte Denkweisen umzustellen sind. Die klassische Physik fordert immer wieder die Entscheidung, einem Objekt die Eigenschaften der Wellen *oder* der Teilchen zuzuordnen. Der von DIRAC vorgeschlagene Versuch ist ein Experiment, das die Unmöglichkeit beweist, mit nur einem Eigenschaftsbild auskommen zu können. In Abb. 5.5.4–1 ist die Versuchsanordnung skizziert. Ein von der Lichtquelle ausgehender Lichtstrahl wird mittels eines halbdurchlässigen Spiegels in zwei Teilstrahlen gleicher Intensität zerlegt. Im eigentlichen Versuch wird die Lichtintensität so weit verringert, daß mit großer Sicherheit nur noch von-

einander getrennte, einzelne Lichtquanten (Photonen) auf den Strahlteiler treffen. In der Versuchsanordnung *a* registriert dann entweder die Fotozelle 1 oder die Fotozelle 2 ein Photon. Die Fotozellen sind auch empfindlich genug, Photonen der halben Energie zu registrieren, aber nie sprechen gleichzeitig beide Fotozellen an. Lichtquanten werden zu 50 % durch den Strahlteiler durchgelassen bzw. über den Umweg geführt. Bringt man in einem zweiten Versuch die Wege der Teilstrahlen wieder zusammen (Teil *b* der Abb.), so ist mit einer Fotozelle keine Aussage über den Weg, den die vereinzelten Quanten gegangen sind, mehr möglich. Eine Fotoplatte registriert aber Interferenzerscheinungen. Einzelne Photonen müssen mit sich selbst interferieren. Das Photon geht – entsprechend unserer Vorstellung von Wellen – beide Wege. Interferenzeffekte sind aber nur zu beobachten, solange die Differenz der Wellenlängen beider Teilwege nicht zu groß wird. Überschreitet sie die Kohärenzlänge (vgl. 2.5.2.3.) – ein Maß für die Längenausdehnung des Photons – interferiert das Photon nicht mehr mit sich selbst. Beide Teilexperimente zeigen den Widerspruch auf, in den man gerät, wenn man entweder das Teilchen oder die Welle sucht. Das Lichtquant vereint in sich beides; jede Eigenschaft ist für sich nachweisbar. Beide Eigenschaften gleichzeitig nachweisen zu können, erscheint nach allen bisherigen Versuchen und theoretischen Betrachtungen unmöglich, beide Versuche stören sich gegenseitig.

Derartige Experimente haben an Wert eingebüßt, da die Fragestellung entweder Welle oder Teilchen seinen Sinn verloren hat. Die Quantentheorie spricht den Grundbausteinen der Materie beide Aspekte zu. In der klassischen Physik wurden frühzeitig beide Aspekte streng getrennt, für sie getrennte Betrachtungsweisen, Theorien aufgebaut. Das ist durch die Quantentheorie korrigiert, auch wenn vieles mit den Formeln der klassischen Theorien ausreichend beschreibbar bleibt.

5.5.4.4. Heisenbergs Unbestimmtheitsrelation. Der Welle-Teilchen-Dualismus liefert für elektromagnetische Wellen das Bild von mit Masse behafteten einzelnen Quanten begrenzter Ausdehnung und für die elementaren Teilchen mit Ruhmasse, z.B. für das Elektron, ein Bild, das neben der Punktförmigkeit für den Sitz der Ladung auch eine wellenförmige Ausdehnung zuläßt, mit der beispielsweise das Elektron im Atom seine stationäre Bahn einstellt (vgl. 5.6.1.). In der klassischen Mechanik rechnet man prinzipiell mit beliebiger Genauigkeit, und für die Wellentheorie kann eine beliebig genaue Wellenlänge bzw. Frequenz vorausgesetzt werden.

Die *Unbestimmtheitsrelation*, auch *Unschärfebeziehung* genannt, gibt ein Maß für die prinzipielle, der Quantennatur innewohnende Ungenauigkeit klassischer physikalischer Größen wie Ort, Impuls und Lebensdauer. Damit ist auch die Grenze für die unbedenkliche Anwendbarkeit solcher Größen festgelegt.

Unabhängig von apparativen Unzulänglichkeiten und dem Geschick der Experimentatoren gilt für die Orts- bzw. Impulskomponenten eines Objektes: $\Delta p_x \Delta x \approx \hbar$, $\Delta p_y \Delta y \approx \hbar$, $\Delta p_z \Delta z \approx \hbar$ (Heisenbergsche Unbestimmtheitsrelation, $\hbar = h/2\pi$).

Das Produkt der einem Objekt innewohnenden Ungenauigkeit des Aufenthaltsortes Δx und des Impulses Δp_x ist etwa gleich dem Planckschen Wirkungsquantum dividiert durch 2π.

Man kann auch formulieren, daß es unmöglich ist, gleichzeitig Ort und Impuls eines Teilchens unendlich genau zu bestimmen. Diese Gleichung gilt nicht nur für die x-, y- und z-Koordinaten. Alle Größen der Dimension „Wirkung" sind dieser Unbestimmtheit unterworfen, unterliegen dieser Quantennatur. So ist die Lebensdauer Δt eines Elektrons auf einem bestimmten Niveau im Atom mit der Niveaubreite ΔW verknüpft zu $\Delta W \cdot \Delta t \approx \hbar$.

Das Zeitintervall, in dem man einen Zustand realisiert findet, ist mit der Genauigkeit, mit der der Zustand auf dem Energiemaßstab festgelegt werden kann, auch durch eine Heisenbergsche Unbestimmtheitsrelation gegeben. Die ordentliche Herleitung der Unbestimmtheitsrelation liefert ein Größergleich-Zeichen ($\geqq \hbar/2$), für die mit Δ bezeichnete Unschärfe ist exakterweise noch eine Meßvorschrift zu formulieren.

In Abb. 5.5.4–2 ist eine anschauliche Darstellung der maximal angebbaren Genauigkeit versucht. Kein elementares Quantum, z.B. weder Elektron noch Photon, kann genauer als durch den schraffierten Bereich gekennzeichnet, lokalisiert werden. Lokalisieren bedeutet hier eine Festlegung bezüglich Ort und Impuls. Die in der Form variable, aber in der Größe nicht zu verringernde Fläche hat die Dimension einer Wirkung (Js). Teilchen und Welle sind für sich allein in der Natur nicht realisierte Idealisierungen.

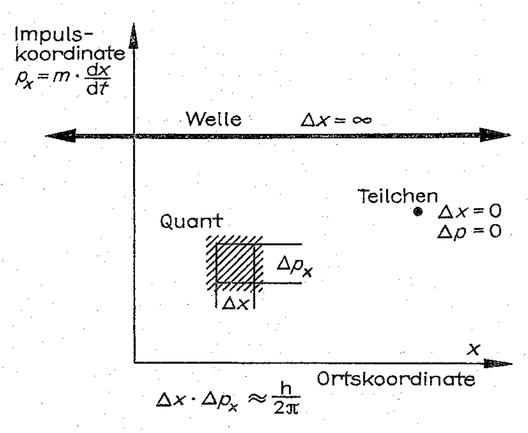


Abb. 5.5.4–2 Heisenbergsche Unbestimmtheitsrelation

5.5.4.5. Wahrscheinlichkeit und Determiniertheit. Der Dirac'sche Spiegelversuch zeigt einen Sachverhalt auf, der dringend eine neue Beschreibungsmöglichkeit erfordert. Wie in Abb. 5.5.4–1 dargestellt, regelt der Strahlteiler das Verhältnis der reflektierten zur durchgelassenen Intensität. Er besteht aus einer teilweise verspiegelten Glasplatte. Je nach Verspiegelungsgrad wird weniger oder mehr durchgelassen, der Rest reflektiert. Das Verhältnis ist genau einstellbar. Wenn nun, wie in dem beschriebenen Experiment, die Strahlintensität so weit verringert wird, daß nur noch einzelne Photonen – beispielsweise je Sekunde eines – auf den Strahlteiler auftreffen, werden nach genügend langer Meßzeit die Mengen der in den beiden Fotozellen registrierten Photonen genau dem vom Strahlteiler vorgegebenen Verhältnis entsprechen. Das einzelne Photon geht aber immer nur einen Weg, entweder zur Fotozelle 1 oder 2, es teilt sich nicht.

So entsteht die Frage, wie sich das Verhältnis einstellt. Betrachtet man nur ein einzelnes Photon, so ist keine sichere Vorhersage möglich, ob es durchgelassen oder reflektiert wird. Es gibt noch andere Experimente, die klar erkennen

lassen, daß es sich nicht um die Unkenntnis über den Ablauf des Ereignisses, das die Wegentscheidung trifft, handelt, sondern daß die im vergangenen Unterabschnitt gegebene Unbestimmtheit als Ursache zu sehen ist.
Im Spiegelexperiment ist der Weg für ein einzelnes Photon nicht exakt vorbestimmt, ist nicht vorhersagbar, nicht determiniert. Die Vorhersagbarkeit auch derartiger Ereignisse war eine Grundlage der exakten Wissenschaften, ein Fundament der klassischen Physik. Es galt, daß, wenn für ein einzelnes Objekt die Bewegungsgesetze und Anfangsbedingungen bekannt waren, der Bewegungsablauf unabdingbar festgelegt, determiniert war. Mit der Entdeckung des Welle-Teilchen-Dualismus und der daraus resultierenden Unbestimmtheitsrelation ergab sich physikalisch wie philosophisch ein neuer Aspekt. Der Weg eines einzelnen Mikroobjektes, ob Photon oder Elektron, ist nicht genau angebbar. Offensichtlich regiert der Zufall den Einzelfall. Eine Vielzahl von Einzelfällen betrachtend, gelangt man wieder zu strengen Gesetzmäßigkeiten. Um dennoch Aussagen über den Einzelfall machen zu können, mußte der Begriff der Wahrscheinlichkeit eines Ereignisses eingeführt werden.
Die Quantentheorie liefert Wahrscheinlichkeitsaussagen für einzelne Meßgrößen, wie Energie, Ort und Impuls. Sie sagt aus, mit welcher Wahrscheinlichkeit sich ein bestimmter Meßwert einstellt. Mit ihr sind die Wahrscheinlichkeitswerte determiniert, bei vorgegebenen Gesetzen und Anfangsbedingungen für bestimmte Ereignisse vorhersagbar. Beispielsweise entspricht der Kreisbahn des Elektrons im Atom als Ergebnis quantenmechanischer Rechnungen einem Maximalwert der Aufenthaltswahrscheinlichkeit. Die ursprüngliche Bahn ist „verschmiert" mit zu kleineren und größeren Radienwerten hin abnehmender Aufenthaltswahrscheinlichkeit. Wahrscheinlichkeitsverteilungen für den Einzelfall entsprechen den Intensitätsverteilungen bei großen Mengen.
Die Quantentheorie enthält als Grenzfall die klassische Physik. Der abstraktere Apparat der Quantentheorie ist immer dann anzuwenden, wenn von der Aufgabenstellung her das Plancksche Wirkungsquantum eine Rolle spielt. Das trifft nicht zu bei Bewegungsabläufen makroskopischer Körper, wohl aber zur Erklärung der Elektronenbewegung im Atom.

5.5.5. Körperstrahlung

In diesem Kapitel werden die Ursachen der Entstehung und der Auflösung elektromagnetischer Wellen behandelt. Unabhängig von der Wellenlänge entsprechen Emission und Absorption von Wärme-, Licht- oder Röntgen-Strahlen der Erzeugung und Vernichtung von Strahlungsquanten. Erzeugung und Vernichtung sind Akte der Wechselwirkung mit Materie, i. allg. mit Atomen oder atomaren Gebilden, wie dem Festkörper. Ohne genaue Kenntnisse des Aufbaues der Atome und seiner Bindungen sind Emission und Absorption nicht zu verstehen.
Erfahrungsgemäß strahlt ein heißer Körper Energie ab, nimmt ein kalter Körper Energie auf. Heiß und kalt sind Relativwerte und beispielsweise bezogen auf einen Mittelwert, der sich, wenn ein Temperaturausgleich erfolgt, einstellen kann. Abgesehen von der Wärmeleitung (in der Beschränkung auf Körper im Vakuum) erfolgt der Energietransport durch die Strahlungsquanten. Sie bewegen sich mit Lichtgeschwindigkeit (c). Entsprechend der Energie eines Quants W (Quant) $= h\nu$; $c = \nu \cdot \lambda$, bedeuten hohe Frequenzen oder kurze Wellenlängen einen großen Energietransport und niedrige Frequenzen, denen lange Wellenlängen entsprechen, einen geringen Energietransport je Elementarquantum der Strahlung. Die elektromagnetischen Wellen der klassischen Physik werden erst durch sehr viele Quanten realisiert. Das menschliche Auge kann, wenn es sich über längere Zeit an Dunkelheit gewöhnt hat, einzelne Lichtquanten wahrnehmen. Die Wärmerezeptoren des Körpers sind nicht so empfindlich. Für Röntgen-Strahlen haben wir keine Empfindungen.
Charakteristisch ist nun, daß die Atome nur Quanten genau bestimmter Frequenzen abstrahlen bzw. absorbieren. Die Energie eines Quants entspricht der Differenz des Energieinhalts eines Atoms nach und vor der Abstrahlung. Folglich entstehen die genau bestimmten Frequenzen durch genau bestimmte Anregungsstufen der Atome. Auch diese Energiequantelung machte die Entwicklung des Apparates der Quantentheorie notwendig und wird erst durch sie erklärt. Alle gebundenen Zustände sind gequantelt. Gleiche Atome haben gleiche Anregungsstufen und strahlen deshalb genau die gleichen Frequenzen ab bzw. absorbieren gleich. Verschiedene Atome haben verschiedene Anregungsstufen. Komplizierte Verbindungen der Atome haben derart viele Anregungsmöglichkeiten, daß die einzelnen Energiestufen ineinander übergehen, ein kontinuierliches Absorptions- bzw. Emissionsband bilden. Jeder Körper ist bestrebt, den niedrigstmöglichen Energiezustand einzunehmen. Wie

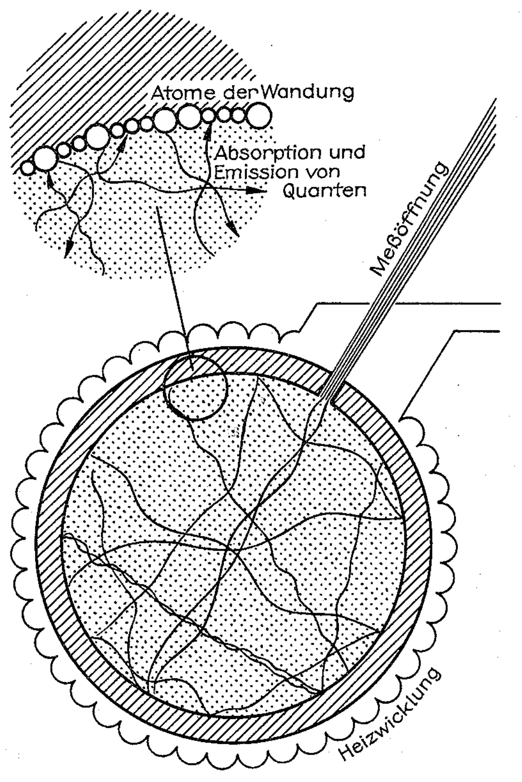


Abb. 5.5.5–1 Schematischer Aufbau eines Hohlraumstrahlers

auch immer Anregungen von Atomen, Molekülen oder Gitterbausteinen erfolgt sein mögen, die Rückkehr in den Grundzustand geschieht von sich aus (spontan) durch Energieabgabe. Daß sich ein Körper i. allg. nicht im energetisch tiefsten, im Grundzustand, befindet, ist damit begründet, daß er ständig von einem Strahlungsfeld umgeben ist, aus dem er Energie aufnimmt. Die Energieaufnahme ist zwangsläufig (induziert, stimuliert). Die Quantenabstrahlung und -aufnahme kommt dadurch ins Gleichgewicht, daß in der Zeiteinheit gleichviel Energie emittiert wie absorbiert wird. Der Energiezustand kann durch die Temperatur des Systems gekennzeichnet werden.

5.5.5.1. Strahlung des schwarzen Körpers (*Hohlraumstrahlung*). Ein schwarzer Körper absorbiert definitionsgemäß alle auftreffende Strahlung, sein Absorptionsvermögen ist 1. Es gibt kein Material, das diesem Idealfall entspricht. Um ihn möglichst gut anzunähern, wird der Hohlraumstrahler verwendet. Abb. 5.5.5–1 zeigt seinen Aufbau. Die Wandung eines Hohlraumes wird von außen durch Heizen auf verschiedene Temperaturen eingestellt. Aus dem Innenraum kann keine Strahlung entweichen. Alle Strahlung wird absorbiert, wie oft sie vorher auch reflektiert werden mag. Das System Strahlenfeld und Wandung im Innern ist nach einiger Zeit im Gleichgewicht, die innere Wandung emittiert und absorbiert gleichviel Energie. Durch ein kleines Loch entweicht etwas Strahlung, unerheblich in der Menge, jedoch ausreichend für Messungen. Diese Strahlung, identisch mit der eines schwarzen Körpers, wird als *Hohlraumstrahlung* bezeichnet.

In Abb. 5.5.5–2 sind als durchgezogene Kurven die frequenzabhängigen Intensitätsverteilungen der Strahlung des schwarzen Körpers aufgezeichnet. Kurven für drei verschiedene Strahlertemperaturen sind dargestellt. Bei niederen Temperaturen wird der Strahler entsprechend dem Maximum der Verteilung rot leuchten und bei 5000 K eine Spektralverteilung haben, die ihn weiß erscheinen läßt. Max PLANCK gelang 1900 die erste befriedigende Beschreibung des Kurvenverlaufes. Die Interpretation seiner exakt die Meßwerte wiedergebenden Verteilungsfunktion der Energiedichte ω_ν kann zugleich als Geburtsstunde der Quantenphysik gelten, $\omega_\nu = 8\pi h\nu^3/c^3(e^{h\nu/kT} - 1)^{-1}$ (*Plancksches Strahlungsgesetz*, ω_ν = Energiedichte, kT = mittlere Energie des Strahlungsfeldes bzw. des Strahlers entsprechend der Temperatur T, c = Lichtgeschwindigkeit).

Zur Interpretation der Gleichung muß man radikal mit grundlegenden Vorstellungen der klassischen Physik brechen.

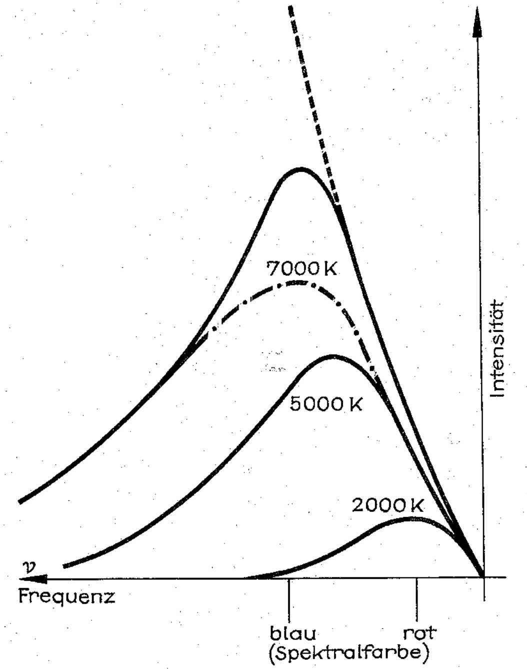


Abb. 5.5.5–2 Strahlung des schwarzen Körpers für verschiedene Temperaturen. Theoretische Beschreibung (7000 K) nach PLANCK ——, nach RAYLEIGH-JEANS – – –, nach WIEN —·—·—

5.5.5.2. Diskussion des Planckschen Strahlungsgesetzes. In der klassischen Physik erklärt man das Entstehen der elektromagnetischen Wellen durch die Energieabstrahlung schwingender Systeme. Die Atome waren dementsprechend Strahlungsoszillatoren, und da sie in der Anregung keinen Auswahlregeln unterlagen, in allen Frequenzen schwingungsfähig. RAYLEIGH und JEANS beschränkten die Zahl der im Hohlraum möglichen Oszillationen. Es sollten nur die Frequenzen entstehen können, die durch stehende Wellen in einem vorgegebenen Volumen gekennzeichnet waren. Die Zahl N der in einem Volumen V möglichen stehenden Wellen berechnet sich in Abhängigkeit von der Frequenz ν für das Frequenzintervall $d\nu$ zu $N_\nu\, d\nu = 8\pi V\nu^2/c^3 d\nu$. Diesen möglichen Schwingungsfreiheitsgraden ist eine mittlere Energie vom Betrage kT (k = Boltzmann-Konstante) zuzuordnen. Damit ergibt sich die frequenzabhängige Energiedichte des Feldes zu $\omega_\nu d\nu = (1/c^3)8\pi\nu^2 kT d\nu$ (*Rayleigh-Jeanssches Strahlungsgesetz*).

Die Formel führt zu dem gestrichelt gezeichneten Kurvenverlauf in der Abb. 5.5.5–2. Das Ergebnis ist unakzeptabel, wenn auch für niedrige Frequenzen richtige Werte entstehen. Das Gesetz enthält den Widerspruch, daß die Energiedichte quadratisch mit ν anwächst und für $\nu \to \infty$ auch gegen Unendlich geht. Das Gesetz war offensichtlich falsch, aber der Fehler in den Voraussetzungen nicht zu erkennen. Dieser unbefriedigende Zustand wurde damals als „Ultraviolettkatastrophe“ bezeichnet und ist unter diesem Begriff, das Versagen der klassischen Physik anzeigend, in die Literatur eingegangen. Nur zwei Voraussetzungen lagen zugrunde. Zum einen die Festlegung über die Zahl der verschiedenen Oszillatoren, zum anderen die Bestimmung des Energiebetrages W $= kT$ je Oszillation.

PLANCK wählte gemäß dem Boltzmannschen Verteilungsgesetz die mittlere Energie der einzelnen Oszillatoren zu $\bar{W} = h\nu(eh\nu/[kT_{-1}])^{-1}$ und erhielt in exzellenter Übereinstimmung mit dem Experiment das nach ihm benannte Strahlungsgesetz.

Das Plancksche Strahlungsgesetz enthält alle wesentlichen, bis zu seiner Entstehung aufgestellten Näherungen. Das Rayleigh-Jeanssche Strahlungsgesetz als Grenzfall für geringe Frequenzen, entsteht durch die Näherung $h\nu/kT \ll 1$ (mittels $eh\nu/kT \approx 1 + h\nu/kT$). Auch ein Versuch der Anpassung an die experimentellen Werte ohne verständliche Interpretation, das Wiensche Gesetz, mit guter Übereinstimmung für hohe Frequenzen, entsteht im Grenzfall $h\nu/kT \gg 1$ (mittels $eh\nu/kT_{-1} \approx eh\nu/kT$).

Ermittelt man die Maxima der spektralen Verteilungen durch Differenzieren nach der Frequenz und Nullsetzen der Ableitung, so ergibt sich das Wiensche Verschiebungs-

gesetz für die Lage der Maxima der Verteilungen in Abhängigkeit von der Temperatur des Systems. $\lambda_{max} \cdot T = 2{,}90 \cdot 10^{-3}$[m · K] (*Wiensches Verschiebungsgesetz*).

Das Stefan-Boltzmann-Gesetz für die gesamte Abstrahlung des schwarzen Strahlers ergibt sich durch die Integration über alle Frequenzen: $u = \sigma \cdot T^4$ (*Stefan-Boltzmannsches Strahlungsgesetz* $\sigma = 5{,}67 \cdot 10^{-8}\,\mathrm{Wm^{-2}K^{-4}}$). Es besagt, daß die totale Energieflußdichte u proportional mit der vierten Potenz der Temperatur ansteigt.

Plancks Strahlungsgesetz nach Einstein abgeleitet. Von Einstein stammt eine sehr interessante und leicht verständliche Herleitung der Planckschen Strahlungsformel. Sie geht von einer Betrachtung der Übergänge zwischen zwei Atomniveaus aus und verwendet die gesicherten Erkenntnisse aus den Frequenzbereichen, wo die Näherungen der klassischen Physik für die Strahlungsdichte mit dem Experiment übereinstimmen (Rayleigh-Jeanssches Strahlungsgesetz). Zum Verständnis der Herleitung ist die Kenntnis der Begriffe, die zur Erklärung des Laserprinzipes (vgl. 5.6.5.4.) eingeführt wurden, notwendig.

Betrachtet seien zwei Atomniveaus nach Abb. 5.5.5–3. Das untere Niveau (mit dem Index 1) kennzeichnet den stabileren Zustand. Die Zahl der Elektronen in den Zuständen seien N_1 bzw. N_2, die Zahl der strahlenden Übergänge in der Zeiteinheit Z_{12} für Übergänge von 1 nach 2, bzw. Z_{21} für Übergänge von 2 nach 1. Im thermodynamischen Gleichgewicht ist die Anzahl der Atome mit besetzten Energiezuständen N_1 bzw. N_2 durch eine Maxwell-Verteilung gegeben: $N_1 = c \cdot \exp[-W_1/kT]$; $N_2 = c \cdot \exp[-W_2/kT]$ (T = Temperatur des Systems, kT = seine mittlere Energie je Atom). Damit schreibt sich das Verhältnis der Besetzungszahlen $N_2/N_1 = \exp[-(W_2 - W_1)/kT]$ bzw. mit $W_2 - W_1 = h\nu$, $N_1/N_2 = \exp(h\nu/kT)$.

Eine andere Darstellung des Verhältnisses der Besetzungszahlen folgt aus Betrachtungen zu den Übergängen zwischen beiden Niveaus. Übergänge von 2 nach 1 erfolgen spontan, von sich aus, ohne äußere Einflüsse. Sie sind zudem noch unabhängig von Einflüssen wie etwa der Existenz eines Strahlungsfeldes. Die Zahl der strahlenden Übergänge in der Zeiteinheit Z_{21} ist proportional der Zahl der Elektronen im Ausgangszustand N_2: Z_{21} (spontan) $\sim N_2$. Außerdem werden noch durch ein äußeres Strahlungsfeld der Frequenz ν Übergänge erzwungen. Diese induzierten Übergänge sind proportional N_2 und der Strahlungsdichte $\varrho(\nu)$, Z_{21} (induziert) $\sim \varrho(\nu) N_2$. Die Gesamtzahl der Übergänge ist also $Z_{21} = [A_{21} + B_{21}\varrho(\nu)]N_2$. A_{21} bzw. B_{21} sind die *Einstein-Koeffizienten* der spontanen bzw. induzierten Emission.

Übergänge unter Absorption von Strahlung, also von 1 nach 2, erfolgen nicht spontan. Zur Anregung der Atome muß Energie zugeführt werden, der Vorgang ist immer erzwungen. Es gilt analog $Z_{12} \sim \varrho(\nu)N_1$; $Z_{12} = B_{12}\varrho(\nu)N_1$. Im Strahlungsgleichgewicht, im Gleichgewicht von Emission und Absorption der Atome der Wandung mit dem Strahlungsfeld gilt $Z_{12} = Z_{21}$ (Hohlraumstrahlung), woraus folgt, daß $N_1/N_2 = [A_{21} + B_{21}\varrho(\nu)]/B_{12}\varrho(\nu)$ wird. Durch Gleichsetzung dieses Verhältnisses der Besetzungszahlen mit dem ersteren ergibt sich die Energiedichte der Strahlung $\varrho(\nu)$ zu $\varrho(\nu) = \dfrac{A_{21}/B_{12}}{\exp(h\nu/kT) - B_{21}/B_{12}}$. Die Bestimmung der Einstein-Koeffizienten erfolgt gemäß dem Bohrschen Korrespondenzprinzip. Die Formel für $\varrho(\nu)$ muß im gesamten Bereich der Frequenz ν gelten, auch im Grenzübergang zur klassischen Physik. Der niederenergetischen Ausläufer der Strahlung des schwarzen Körpers $h\nu \ll kT$ ist gut durch das Rayleigh-Jeanssche Strahlungsgesetz beschrieben $\varrho(\nu) = kT 8\pi\nu^2/c^3$.

Einsteins Darstellung der Strahlungsdichte geht nur für $B_{21} = B_{12}$ und $A_{21}/B_{12} = h\nu 8\pi\nu^2/c^3$ in das Rayleigh-Jeanssche Strahlungsgesetz über, so daß derart die Koeffzienten zu bestimmen sind. $B_{12} = B_{21}$ besagt, daß die Übergangswahrscheinlichkeiten für die induzierte Emission und induzierte Absorption gleich sind. $A_{21} = B_{21} \cdot 8\pi h\nu^3/c^3$ zeigt die Proportionalität zwischen den Übergangswahrscheinlichkeiten für die spontane und induzierte Emission für gleiche Frequenzen an. Die spontane Emission wächst mit ν^3 relativ zur induzierten Emission. Durch diese Koeffizientenbestimmung ist das Plancksche Strahlungsgesetz

$$\varrho(\nu) = \frac{8\pi h\nu^3}{c^3} \, \frac{1}{e^{h\nu/kT} - 1}$$

wie schon angegeben (vgl. 5.5.5.2.) abgeleitet.

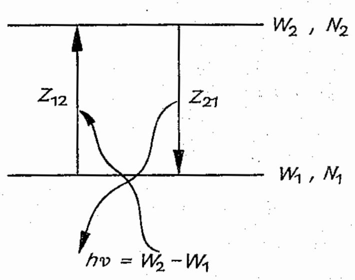


Abb. 5.5.5–3 Zwei Atomniveaus der Energie W_1 bzw. W_2, mit den Besetzungszahlen N_1 bzw. N_2 und der Zahl der strahlenden Übergänge in der Zeiteinheit Z_{12} bzw. Z_{21}

5.5.5.3. Das Plancksche Wirkungsquantum. Das Plancksche Strahlungsgesetz war bald als das richtige akzeptiert. Die Interpretation vertiefte sich im Laufe der Jahre. Es kamen grundlegende Hypothesen hinzu, wie die Einsteinschen Lichtquanten und das Bohrsche Atommodell. Mit der durchgängigen Formulierung der Quantentheorie, gut zwanzig Jahre nach Plancks Entdeckung, war es dann vollständig fundiert.

Aus den experimentellen Werten bestimmte Planck schon die Größe von h zu $6{,}55 \cdot 10^{-34}$ Js. Der genaue Wert beträgt $h = 6{,}626\,176 \cdot 10^{-34}$ Js. „Die Einführung der Konstanten bedeutete einen Bruch mit der klassischen Theorie, der viel radikaler war, als ich anfänglich vermutet hatte", schrieb Planck viele Jahre später.

Durch die Konstante wurden die Energien der Oszillatoren quantisiert, auf Werte festgelegt, die sich nur um h ändern konnten. Die klassische Physik ordnete den Oszillatoren, gleich schwingender Pendel, eine von der maximalen Auslenkung abhängige Energie zu. Die Auslenkung war, wie die Amplituden der elektromagnetischen Wellen, kontinuierlich einstellbar. Planck brachte Diskontinuität in die bis dato „natürliche" Stetigkeit. Die Quantisierung der Oszillatoren und damit des Strahlungsfeldes hatte zur Folge, daß neben der Existenz der Energiestufen sprunghafte Änderungen eingeführt werden mußten. Von einer Energiestufe zur anderen gelangt man nur über Sprünge, über diskontinuierliche Änderungen. Einsteins Photonen, ihre Emission und Absorption, gaben dafür eine gute Verständnismöglichkeit. Das Plancksche Wirkungsquantum ist heute als eine universelle Naturkonstante verstanden, die die Grenze aller Beschreibungen im klassischen Sinne festlegt.

5.6. Atombau und Spektrallinien

Die Überschrift stellt, wie der Titel eines Buches von SOMMERFELD, den Zusammenhang zwischen Stoff und Strahlung her. Präzise Informationen über den Aufbau der Atome liefern die abgestrahlten und gleichermaßen die absorbierten elektromagnetischen Wellen. Atomzustandsänderungen sind Energieänderungen. Die Energiedifferenzen können dem Feld entnommen werden, bzw. durch Abstrahlung der Atome in Erscheinung treten.

Verdünnte Gase können durch ein elektrisches Feld zum Leuchten angeregt werden. Dieses Leuchten wurde frühzeitig analysiert. Dabei fand man – wie leicht mit einem Prisma nachvollziehbar – daß es meistens nur aus einzelnen Spektrallinien besteht. Es zeigte sich, daß jedem Element ein Linienspektrum mit charakteristischer Linienfolge eigen ist. Wasserstoffatome liefern das einfachste Spektrum. Ihre sechs Linien im sichtbaren Bereich waren so regelmäßig angeordnet, daß schon im Jahre 1885 BALMER eine Formel für die Frequenzen der einzelnen Linien aufstellen konnte. Er fand empirisch einen einfachen Zusammenhang zwischen den Frequenzen der Linien des Wasserstoffs, der auch alle weiteren später festgestellten Linien der Serie im nichtsichtbaren Teil des Spektrums mit erfaßte.

BALMERS Serienformel, umgerechnet auf die Frequenzen ν_n des ausgesandten Lichtes, lautet: $\nu_n = R(1/4 - 1/n^2)$, *Balmer-Serie* (R = Rydberg-Konstante $= 3{,}29 \cdot 10^{15}\,\mathrm{s}^{-1}$).

Für n sind ganze Zahlen $n \geqq 3$ einzusetzen. Für $n = 3$ beispielsweise ergibt sich die rote Wasserstofflinie mit $\nu = 4{,}57 \cdot 10^{14}\,\mathrm{s}^{-1}$ entsprechend einer Wellenlänge ($\lambda = c/\nu$) $\lambda_3 = 656\,\mathrm{nm}$.

Auch die Verallgemeinerung obiger Formel zu $\nu_{nn'} = R(1/n'^2 - 1/n^2)$, die alle weiteren, später gefundenen Serien des Wasserstoffs beschreibt, stammt noch von BALMER.

Damals war man noch weit entfernt von brauchbaren Atommodellen, aber es wurde klar, daß jede Modellvorstellung über das Atom an der Erklärung über das Zustandekommen der Spektrallinien geprüft werden würde.

RUTHERFORD schloß aus seinen Streuversuchen (vgl. 5.5.2.) auf einen dem Sonnensystem ähnlichen Aufbau der Atome. Nach seinem *Planetenmodell* umkreisen die Elektronen den sehr schweren, fast ruhenden, positiv geladenen Kern. Ihre Kreisbahnen sind durch das Gleichgewicht zwischen der anziehenden Coulomb-Kraft und der Abstoßung durch die Zentrifugalkraft bestimmt (Abb. 5.6.1–1): Zentralkraft = Fliehkraft, $Ze^2/(4\pi\varepsilon_0 r^2) = mr\omega^2$ (ω = Winkelgeschwindigkeit, m = Masse der Elektronen, r = Bahnradius, Ze = Kernladung, $\varepsilon_0 = 8{,}86 \cdot 10^{-12}\,\mathrm{AsV^{-1}m^{-1}}$ = Dielektrizitätskonstante des Vakuums).

Das ist der einfachste Planetenmodellansatz, wobei 1. der Atomkern als in Ruhe bleibend betrachtet wird. Eine Kernmitbewegung kann durch eine Korrektur nachträglich berücksichtigt werden. 2. erfaßt die Beschreibung auf Kreisbahnen nicht alle möglichen Fälle. Wie in der Himmelsmechanik sind Ellipsenbahnen der allgemeinere Fall. 3. wird erst einmal nichtrelativistisch gerechnet und dabei die Abhängigkeit der Elektronenmasse von der Geschwindigkeit vernachlässigt.

RUTHERFORDS Atommodell ist das erste mit einer richtigen wissenschaftlichen Vorstellung. Der Planetenmodellansatz hat sich neben seiner Anschaulichkeit auch als sehr tragfähig erwiesen. Aber dieses Modell erklärt nicht die Spektrallinien, weder die des Wasserstoffs noch überhaupt bestimmte Frequenzen. Die Elektronen können je nach ihrer Geschwindigkeit jeden Bahnradius, jeden Energiewert annehmen. Energieänderungen sind auch mit jedem Betrag möglich, so daß das abgestrahlte Licht alle Frequenzen enthalten, ein kontinuierliches Spektrum liefern müßte. Unverständlich ist die Stabilität des Atoms überhaupt. Nach der klassischen Elektrodynamik strahlt eine beschleunigte Ladung, also auch das sich im Kreise bewegende Elektron, beständig Energie ab. Das Atom müßte als elektrischer Dipol wirken, und das Elektron, durch die Energieabgabe mit immer kleiner werdendem Bahnradius kreisend, nach kurzer Zeit auf den Kern treffen.

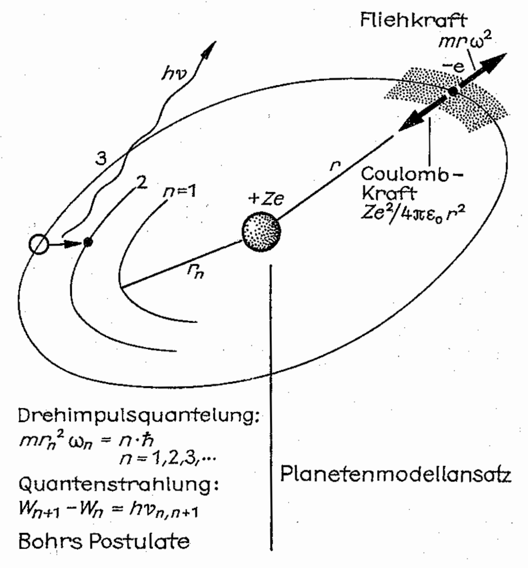


Abb. 5.6.1–1 Bohrsches Atommodell

5.6.1. Das Atommodell von BOHR

5.6.1.1. BOHRS Postulate. BOHR folgerte aus Betrachtungen zu den damaligen Vorstellungen über das Atom, daß das Plancksche Wirkungsquantum in die Theorie eingeführt werden müßte. Es ging um das Fehlen einer Länge bei den gegebenen Grundgrößen e und m im Rutherfordschen Modell. Die Elementarlänge wäre aber über $h^2\varepsilon_0/(me^2)$ dimensionsmäßig einführbar. Er ging davon aus, daß das Rutherfordsche Atommodell die richtige Basis zur Berechnung der Spektren wäre, daß aber die Erzeugung einer Spektrallinie ein Quantenprozeß sein müßte.

BOHR formulierte 1913 Forderungen – Voraussetzungen, die zwar nicht beweisbar, aber notwendig zur Erklärung der Spektrallinie waren (*Bohrsche Postulate*):

- Die Elektronen im Atom befinden sich auf genau festgelegten stationären Umlaufbahnen. Die Bahnen sind dadurch festgelegt, daß der Drehimpuls gequantelt wird, d. h. nur ganzzahlige Vielfache von $\hbar = h/(2\pi)$ betragen darf. Elektronen auf diesen stationären Umlaufbahnen strahlen keine Energie ab.
- Die Elektronen können ihre Bahnen wechseln. Beim Übergang in eine kernnähere Bahn wird ein Energiequant in Form von elektromagnetischer Strahlung abgegeben.

Abb. 5.6.1–1 skizziert den Sachverhalt. Aus dem Planetenmodellansatz nach RUTHERFORD entstehen durch die Drehimpulsquantelung Bahnen mit nur noch bestimmten erlaubten Radien $r_n = 4\pi\varepsilon_0\hbar^2/(mZe^2)\cdot n^2$; $n = 1, 2, 3, \ldots$
Die Radien wachsen quadratisch mit n. Für den einfachsten Fall, das Wasserstoffatom mit der Kernladungszahl $Z = 1$, ergibt sich der kleinste Radius ($n = 1$) zu $r_1(H) = 4\pi\varepsilon_0\hbar^2/(me^2) = a_H = 0{,}53\cdot 10^{-10}$ m (a_H = *Bohrscher Radius*). Mit ihm schreibt sich die vorherige Gleichung für den Radius der n-ten Bahn eines Atoms der Kernladungszahl Z nun zu $r_n = a_H n^2/Z$. Die Bahnenergie der Elektronen errechnet sich aus der potentiellen und der kinetischen Energie der Elektronen. Die potentielle Energie ist gleich der Arbeit, die notwendig ist, das Elektron aus dem Unendlichen auf den Bahnradius r zu bringen [$W_n^{pot} = -Ze^2/4\pi\varepsilon_0 r_n$)]. Die kinetische Energie ergibt sich zu $-1/2\ W_n^{pot}$, und damit wird die Bahnenergie $W_n = -\hbar RZ^2 1/n^2$; $R = me^4/(8\varepsilon_0^2 h^3)$.
Daß der Wert negativ ist, entspricht der Konvention, Energie der Bindung negativ anzusetzen, um die kinetische Energie der ungebundenen, der freien Elektronen bzw. des abgestrahlten Lichtes positiv zu erhalten. Beim Übergang von der Bahn n' zur Bahn $n (n' > n)$ ist die Energiedifferenz $W_n - W_{n'} = \hbar R/Z^2\cdot(1/n'^2 - 1/n^2)$. Also ergibt sich die Frequenz der ausgesandten Strahlung gemäß $h\nu = W_n - W_{n'}$ zu $\nu = RZ^2(1/n'^2 - 1/n^2)$, in Übereinstimmung mit BALMERS empirisch gefundener Formel. Die Rydberg-Konstante ist nun nicht wie bei BALMER den beobachteten Spektralwerten angepaßt, sondern auf Naturkonstanten zurückgeführt. Es bestätigte sich der damalige experimentelle Wert.

Bohrsches Korrespondenzprinzip. BOHR formulierte noch ein drittes Postulat, was auch als Bohrsches Korrespondenzprinzip bekannt ist:

- Die Gesetze der neuen Atommechanik müssen für große Quantenzahlen n in die entsprechenden Gesetze der klassischen Physik übergehen.

Für große n werden die Differenzen zwischen den Energiestufen immer geringer. Die Energiestufen liegen dicht an dicht und gehen im Grenzfall für $n \to \infty$ in ein Kontinuum über – dort sollte die klassische Physik gelten. Im Grenzfall ergibt die Serienformel: $\lim_{n,\, n\to\infty} \nu = \nu_\infty$ $\nu_\infty = RZ^2 2(n' - n)/n^3$.
Gilt nun die klassische Elektrodynamik, so kann nur $n' - n = 1$ sein, da dort nur die Grundfrequenz, die Dipolstrahlung, auftritt. Setzt man R ein, wird

$\nu_\infty = [Ze^2/(4\pi\varepsilon_0)^2\cdot m/(2\pi(n\hbar)^3]$.

Der Planetenansatz lieferte $Ze^2/(4\pi\varepsilon_0) = mr^3\omega^2$, so daß sich ergibt $(n\hbar)^3 = m^3r^6\omega^4/(2\pi\varepsilon_\infty)$. Unter der Bedingung, daß die Frequenz der Strahlung gleich der Umlauffrequenz des Elektrons ist ($\nu_\infty = \omega/2\pi$), und das ist im Falle der Dipolstrahlung so, ergibt sich die Quantelung des Drehimpulses $n\hbar = mr^2\omega$. Die Drehimpulsquantelung und die gequantelte Energieabstrahlung ergeben also im Grenzfall die Dipolstrahlung der klassischen Elektrodynamik. Man kann auch umgekehrt – wie schon BOHR es ausführte – formulieren, daß die gequantelte Energiestrahlung unter der Voraussetzung, daß im Grenzfall die klassische Elektrodynamik gelten soll, die Drehimpulsquantelung begründet.

5.6.1.2. Franck-Hertz-Versuch. James FRANCK und Gustav HERTZ zeigten 1914, daß bei Stößen von Elektronen mit Molekülen von Edelgasen und Metalldämpfen die Elektronen, solange ihre kinetische Energie eine bestimmte Größe nicht übersteigt, nur elastisch reflektiert werden. Sobald die kinetische Energie diese bestimmte Größe erreicht, verliert das Elektron seine gesamte kinetische Energie an das Atom. Das entscheidende Experiment wurde mit Hg-Dampf durchgeführt. Der charakteristische Energiewert entsprach (in der Umrechnung mit W_{kin} (Elektron) $= h\cdot\nu_{Hg}$) genau der Energie der Lichtquanten $h\nu_{Hg}$ einer auch beobachteten Emissionslinie im fernen Ultraviolett. Es war gezeigt, daß durch Elektronenstoß ein Atom auf eine bestimmte Energiestufe gehoben, in einen charakteristischen Anregungszustand gebracht werden konnte. In der energetischen Umgebung existiert nur dieser eine Anregungszustand. Das Atom kehrt unter Aussendung eines Lichtquantes mit der Energie dieser Stufe in den Grundzustand zurück.
Das war ein entscheidendes Experiment, ein grundlegender Versuch zur Quantenphysik. Er bestätigte die Existenz von Energiestufen für die Atomanregung durch Elektronenstoß und damit die gequantelten Energiestufen der Bohrschen Atommodellvorstellungen.
Abb. 5.6.1–2 zeigt den prinzipiellen Versuchsaufbau des Franck-Hertz-Versuches, wie er an Hochschulen und auch als Schülerexperiment heutzutage existiert. Auf den Nachweis der Ultraviolett-Linie mit der Wellenlänge $\lambda = 253{,}6$ nm wird allgemein verzichtet, man begnügt sich mit dem Nachweis der Existenz der Anregungs-Energiestufen der Quecksilberatome. In einer auf etwa 150 °C geheizten Vakuumröhre bildet das eingelagerte Quecksilber Dampf mit $\approx 10^{-2}$ Pa Druck. Mittels der Glühelektrode werden Elektronen erzeugt und durch eine einstellbare Spannung U_B zum Gitterraum bis auf die Energie von $W_{kin} = eU_B$ beschleunigt. Durch den Gitterraum fliegen die Elektronen, hin und wieder auf Quecksilberatome stoßend, in Richtung Anode. Zwischen Gitterraum und Anode ist ein schwaches Sperrpotential gelegt. Im Ergebnis zeigt das Strommeßgerät einen mit der Spannung nicht stetig ansteigenden Strom. Bei bestimmten Spannungswerten fällt der Strom steil ab. Die dort fehlenden Elektronen haben nur noch eine so geringe Energie, daß sie das schwache Sperrpotential nicht überwinden können, sie haben ihre Energie an die Hg-Atome abgegeben. Der durch das Experiment ausgezeichnete Spannungswert von $U_B = 4{,}9$ V entspricht einer Anregung der Quecksilberatome auf eine Energie von $W = 4{,}9$ eV. Die weiteren Minima bei Spannungen von 9,8, 14,7, ... eV entstehen durch nacheinander mehrmalige Energieabgabe eines Elektrons. FRANCK und HERTZ hatten aus dem ausgezeichneten Spannungswert und der Wellenlänge λ_{Hg} der bei der Abregung der Atome emittierten Strahlung die Größe des damals noch nicht experimentell genau ermittelten Wirkungsquantums h bestimmt ($eU_{Hg} = hc/\lambda$Hg).

5.6.1.3. Das Wasserstoffatom. Der Wasserstoff bildet das einfachste der möglichen Atome. Der Kern besteht aus nur einem Proton und die Hülle aus einem einzigen Elektron. Dieses Atom hat die geringste Zahl von Anregungsstufen und demzufolge das einfachste Spektrum.
Aus der von RYDBERG angegebenen und mit der Bohrschen Theorie abgeleiteten Formel für die Frequenzen der Spektrallinien ergeben sich Linienserien, wenn man n' einen festen Wert zuordnet und n alle Werte durchlaufen läßt. Die Serien tragen die Namen ihrer Entdecker, $n' = 1$ = Lyman-Serie, $n' = 2$ = Balmer-Serie, $n' = 3$ = Paschen-Serie, $n' = 4$ = Bracket-Serie, $n' = 5$ = Pfund-Serie. Die Linien der Lyman-Serie liegen auf der ultravioletten Seite neben dem Bereich des sichtbaren Lichtes, die ersten Linien der Balmer-Serie sind zwischen Rot und Violett im Sichtbaren zu finden, während alle weiteren Serien sich auf der infraroten

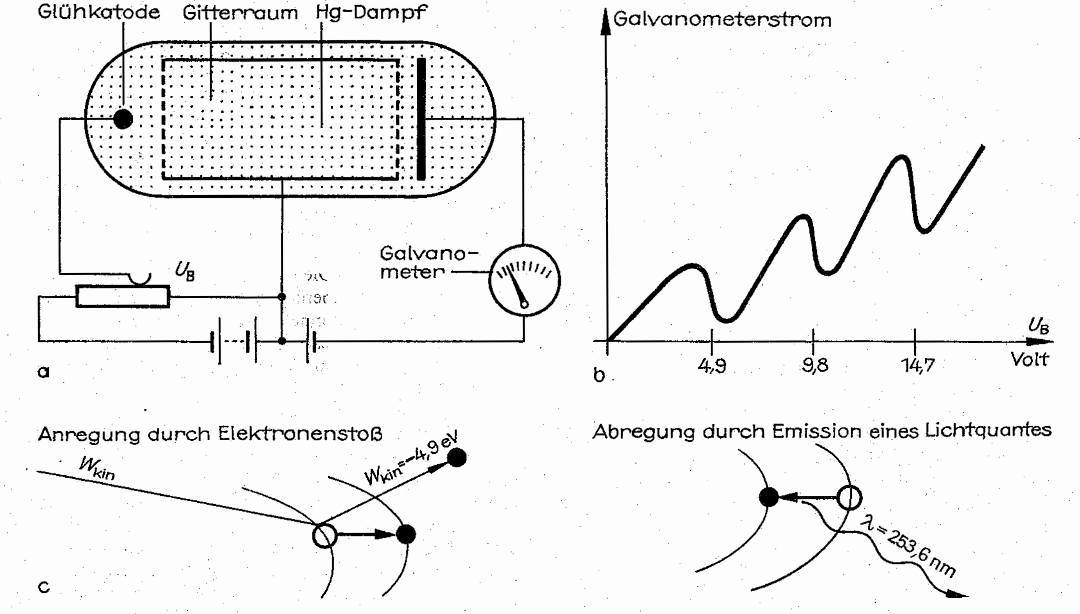


Abb. 5.6.1–2 Franck-Hertz-Versuch, (a) Versuchsaufbau, (b) Versuchsergebnis, (c) Interpretation im Bohrschen Atommodell

Bindungsenergie in eV

0, −1, −2, −3, −4, −5, −10, −13,6

∞, 4, 3, 2, 1

n, n'

Seriengrenze

freie Elektronen

gebundene Elektronen

Absorption-, Emission von Lichtquanten

Balmer-Serie
$n' = 2, n = 3, 4, 5, \ldots$

$$\nu_{n,n'} = R\left(\frac{1}{n'^2} - \frac{1}{n^2}\right)$$

Lyman-Serie
$n' = 1, n = 2, 3, 4 \ldots$

Grundzustand

Abb. 5.6.1–3 Darstellung der Energieniveaus der Atome und der Übergänge am Beispiel des Wasserstoffes

Seite für das menschliche Auge auch nicht sichtbar erstrekken.
Es sind nicht alle Serien in Emission und Absorption gleichermaßen zu beobachten. In Emission – bei der Abstrahlung von Licht – sind die Serien wie hier beschrieben zu beobachten. Jedoch in Absorption, wenn man beispielsweise Wasserstoff mit Licht aller Wellenlängen bestrahlt und beobachtet, welche Frequenzen fehlen, von den Atomen absorbiert wurden, so ist nur die Lyman-Serie zu finden. Normalerweise befinden sich die Atome im Grundzustand, so daß nur Linien, vom Grundzustand ausgehend, auftreten können, sich nur die Lyman-Serie ausbildet.
Um Spektrallinien darstellen zu können, wählt man ein *Termschema* im Energiemaßstab. Abb. 5.6.1–3 zeigt die Energieniveaus der Atome und die Spektrallinien am Beispiel des Wasserstoffes. In diesem Termschema ist, wie allgemein üblich, die Bindungsenergie des Elektrons im Atom aufgetragen. Ein Elektron mit $n = 1$ befindet sich auf der kernnähesten Bahn. Mit wachsendem n wachsen die Abstände vom Kern proportional mit n^2 an und nimmt die Bindungsenergie mit $1/n^2$ ab. Die Frequenzen einer Serie gehen mit $n \to \infty$ gegen einen Grenzwert, die *Seriengrenze*. Diese Grenzfrequenz charakterisiert Elektronen auf Bahnen weit ab vom Kern, mit verschwindend geringer Bindungsenergie. Die Linien liegen energetisch so dicht, daß man vom Seriengrenzkontinuum spricht. Im Seriengrenzkontinuum vollzieht sich der Übergang aus den gebundenen Zuständen in das Kontinuum der freien Elektronen. Dem entsprechend gehört zur Seriengrenze die Bindungsenergie Null.
Reduzierte Massen. In der optischen Spektroskopie kann man sehr genau messen, beispielsweise Spektrallinien auf sieben Stellen genau angeben. Die Genauigkeit der Berechnung ist in der Rydbergschen Formel nur durch die Genauigkeit, mit der die Rydberg-Konstante R gegeben ist, begrenzt. Schon frühzeitig wurde erkannt, daß diese Konstante bereits in der fünften Stelle zu Abweichungen von den experimentellen Werten führte. Eine wesentlich größere Genauigkeit läßt sich bei Berücksichtigung der Mitbewegung des Kerns erreichen. Bisher war angenommen, daß das Elektron um den ruhenden Kern kreist. Diese Annahme ist nur bei unendlich großer Masse des Kernes erfüllt, ansonsten wird er sich ein klein wenig mitbewegen, selbst etwas kinetische Energie besitzen. Das Problem läßt sich rechnen, wenn der Ursprung des Koordinatensystems in den Schwerpunkt auf der Linie zwischen Kern und Elektron gelegt wird. Das Ergebnis kann wieder als eine Bewegungsgleichung in der ursprünglichen Form geschrieben werden, d. h. es erscheinen nicht die Radiusvektoren des Schwerpunktsystems, sondern wieder nur der alte Radiusvektor vom Kern zum Elektron. Dafür ist die ursprüngliche Masse des Elektrons m durch den Ausdruck $\mu = m/(1 + m/M)$ (*reduzierte Masse*) zu ersetzen. M ist die Kernmasse. Damit ist das Zwei-Teilchenproblem auf ein Ein-Teilchenproblem reduziert. Es ist nur die Bewegung von m zu betrachten, die Mitbewegung von M ist durch Einsetzen einer „reduzierten Masse" berücksichtigt.
So ergeben sich beispielsweise die Rydberg-Konstanten für Wasserstoff R_H und Helium R_{He} zu $R_H = R(1 + m/M_H)^{-1}$ bzw. $R_{He} = R(1 + m/M_{He})^{-1}$ (M_H = Masse des Wasserstoffkernes, M_{He} = Masse des Heliumkernes, R = Rydberg-Konstante für den unendlich schweren Kern, $R_H = 3{,}28805 \times 10^{15}\,s^{-1}$, $R_{He} = 3{,}28939 \cdot 10^{15}\,s^{-1}$, $R = 3{,}28984 \cdot 10^{15}\,s^{-1}$.
Selbst Isotopieeffekte lassen sich so berücksichtigen. Die rote Linie der Balmer-Serie für Wasserstoff unterscheidet sich experimentell von der des Deuteriums um ≈0,18 nm in der Wellenlänge. Die Berücksichtigung der Neutronenmasse im Kern des Deuterons zur Berechnung der reduzierten Masse erklärt genau diese Wellenlängendifferenz. Der Unterschied zwischen den beiden Spektrallinien war es, der 1932 zur Entdeckung des Deuteriums führte.
Wasserstoffähnliche Spektren. Das Modell des Wasserstoffatoms gilt auch noch für ionisierte Atome mit nur einem verbleibenden Elektron. Das betrifft das einfach ionisierte Helium (He^+), das zweifach ionisierte Lithium (Li^{++}), das dreifach ionisierte Beryllium (Be^{+++}) usw. Bei diesen Ionen umkreist, wie im Falle des Wasserstoffs, nur ein Elektron den Kern. Die größere Kernmasse wird in der Rydberg-Konstanten R berücksichtigt und die Kernladung Z durch Z^2 in der Formel für die Spektrallinien $\nu_{n,n'} = RZ^2(1/n'^2 - 1/n^2)$. Diese Spektren sind ähnlich denen des Wasserstoffs, nur um Z^2 zum Ultravioletten hin verschoben. Speziell stimmen beim Heliumion für $n' = 4$ (Pikkering-Serie) die zu geraden Werten von n gehörigen Linien mit den Linien der Balmer-Serie fast überein. Die feine Abweichung durch die verschiedenen Massen der Kerne des Wasserstoffs und des Heliums führte dazu, die erst einmal dem Wasserstoff zugeschriebenen Linien der Pickering-Serie als Linien des Heliumions zu erkennen.
5.6.1.4. Sommerfelds Modellerweiterung. Die Spektren des ungestörten Wasserstoffatoms waren mit dem Bohrschen Atommodell ausreichend und sogar verblüffend genau erklärt. Die Spektrallinien aller anderen Atome aber zeigten eine zu große Linienvielfalt, um in so einfacher Weise beschrieben werden zu können. Auch prinzipiell war, wie schon von Bohr erkannt, die Beschränkung in den ersten Rechnungen auf Kreisbahnen nicht gerechtfertigt. Genau wie in der Himmelsmechanik mußte der allgemeinere Fall, mußten Ellipsen als mögliche stationäre Bahnen für die Elektronen im Atom zugelassen werden. Diesen Weg ging mit ausführlichen Berechnungen Sommerfeld 1916. Er verwendete genau wie Bohr die klassische Mechanik des Planetenmodelles und quantisierte es wiederum über den Drehimpuls. Dieses Vorgehen kennzeichnet Rechnungen, die man heute zur „älteren Quantentheorie" zählt. Es war prinzipiell unverstanden, aber vom Ergebnis her richtig, da es gestattete, die gemessenen Spektrallinien in Formeln zu erfassen, vieles zu erklären und weitere Serien vorherzusagen.
Nichtrelativistische Rechnungen ergeben, daß die Energien von Elektronen auf Ellipsenbahnen nur von der großen Halbachse abhängen und unabhängig von der auch durch die kleine Halbachse gegebenen Form sind. Alle Ellipsen mit gleichgroßer Halbachse A führen zur gleichen Bahnenergie, und diese ist gleich der des Kreises mit dem Radius $R = A$. Den Drehimpuls betrachtend, ergibt sich einmal, daß er überall längs der Bahn konstant ist (2. Keplersches Gesetz, Flächensatz) und zum anderen, daß er auch von der kleinen Halbachse abhängt. So existiert die Möglichkeit, besser Notwendigkeit, den Drehimpuls ein zweites Mal zu quanteln (die erste Quantelung legte die Energie und damit die große Halbachse fest). Die Quantelung des Drehimpulses der Ellipsen bei gegebener großen Halbachse A_n beschränkt auch die kleine Halbachse B auf nur diskrete Werte,

$$B_{n,k} = \frac{a_H}{Z} nk,\ k = 1, 2, 3, \ldots, n;\ A_n = \frac{a_H}{Z} n^2,\ n = 1, 2, 3, \ldots$$

n wird Hauptquantenzahl genannt und $l = k - 1$ Nebenquantenzahl bzw. Drehimpulsquantenzahl. Warum man nun l statt k verwendet, ist im folgenden zu erläutern.

SOMMERFELDS Rechnungen unter Berücksichtigung relativistischer Effekte (Abhängigkeit der Masse von der Geschwindigkeit) führten zu folgendem Ausdruck für die Energie der Elektronen auf Ellipsenbahnen:

$$W_{n,l} = W_n \left[1 + a^2 \frac{Z^2}{n^2} \left(\frac{n}{l+1} - \frac{3}{4} \right) + \ldots \right].$$

$n = 1, 2, 3, \ldots$: Hauptquantenzahl, $l = 0, 1, 2, \ldots, n-1$: Drehimpulsquantenzahl, $a = e^2/(4\pi\varepsilon_0 \hbar c)$: *Sommerfeldsche Feinstrukturkonstante*. a entspricht dem Verhältnis der Umlaufgeschwindigkeit eines Elektrons auf der ersten Bohrschen Bahn des Wasserstoffs zur Lichtgeschwindigkeit c und beträgt 1/137.

Die obige Gleichung entsteht mittels einer Reihenentwicklung, wobei nach dem 2. Glied abgebrochen wird. Das erste Glied enthält nur die Energie entsprechend einer Kreisbahn nach BOHR W_n. Das zweite Glied beschreibt zusätzlich, dichtbenachbarte Spektrallinien, die zu Ellipsenbahnen gehören.

Nachdem mit der durchgängigen Formulierung der Quantentheorie durch SCHRÖDINGER und HEISENBERG das Atom exakt fundiert berechnet werden konnte, bestätigten sich zwar die sehr genauen Werte nach BOHR und SOMMERFELD, aber folgendes mußte korrigiert werden: Die Quantelung des Drehimpulses hat in der Form $[\boldsymbol{V} \times \boldsymbol{S}] = \sqrt{l(l+1)}\ \hbar$, $l = 0, 1, 2, \ldots, n-1$ zu erfolgen, und es existiert auch der Drehimpuls 0, die Drehimpulsquantenzahl $l = 0$. Das entspricht einer zur Pendelbahn entarteten Ellipse. Das Elektron pendelt – unvereinbar mit den Vorstellungen der klassischen Physik – genau durch den Kern. Hier offenbaren sich die Grenzen des Bohr-Sommerfeldschen Modellbildes. Wegen seiner guten Anschaulichkeit wird es auch im folgenden verwendet. Notwendige Korrekturen aus den Ergebnissen der Wellenmechanik sind immer angemerkt.

5.6.1.5. Das Leuchtelektronenmodell. Beim Wasserstoff ist die Energie für alle Bahnen einer Hauptquantenzahl n gleich, unabhängig von der Drehimpulsquantenzahl. Das gilt nur für dieses einfachste der Atome. Der feine Unterschied aus relativistischer Rechnung sei hier nicht beachtet. Bei den Alkalien (Li, Na, K, Rb, Cs, Fr) gestattet der Atombau eine auch noch relativ einfache Betrachtungsweise, die auf der des Wasserstoffs aufbaut. Die Alkaliatome haben als Charakteristikum ein einzelnes, auf einer Bahn mit großem Kernabstand kreisendes Elektron. Alle anderen Elektronen sind relativ fest gebunden und reaktionsträge in „abgeschlossenen Schalen" wesentlich näher am Kern gelagert. Diese Atome sind somit gebildet aus dem „Atomrumpf", der aus dem Kern und $Z - 1$ „Rumpfelektronen" besteht, und dem einzelnen äußeren, dem *Leuchtelektron*. Der Name Leuchtelektron entstand, weil die Energiedifferenzen seiner Terme Spektrallinien im sichtbaren Bereich des Lichtes bilden. Mehr noch ist der Begriff Valenzelektron gebräulich, da diese Elektronen die chemische Valenz festlegen. Der einfach positiv geladene Rumpf und das Leuchtelektron bilden ein dem Wasserstoff ähnliches Atom. Die Spektren der Alkaliatome und ihre Serienformeln weisen das aus.

In Abb. 5.6.1–4 ist schematisiert das Spektrum eines Alkaliatoms dargestellt. Zur besseren Übersicht werden die Linien der Serien zu verschiedenen Drehimpulswerten nebeneinander gezeichnet. Die einzelnen Terme sind mit Zahlen und Buchstaben belegt. Die Zahlen bezeichnen den Wert der Hauptquantenzahl n, die Buchstaben stehen für verschiedene l-Werte. Die Bezeichnung ist historisch entstanden (*s*harp-, *p*rincipal-, *d*iffuse- und *f*undamental series). Beispielsweise gehört der 3p-Term zur Serie mit der Drehimpulsquantenzahl $l = 1$, die Hauptquantenzahl beträgt $n = 3$. Im Vergleich zum Wasserstoff, wo die Terme gleicher Hauptquantenzahl zusammenfallen, sind sie hier deutlich unterschieden. Die fehlenden Terme 1s, 2s und 2p gehören zum Atomrumpf. Das Leuchtelektron ist im Grundzustand auf dem 3s-Niveau zu finden. Die Formel für die Frequenzen der Spektrallinien der Serien ähnelt der des Wasserstoffes: $\nu_{n,n'} = R[1/(n'+a)^2 - 1/(n+b)^2]$. Die Korrekturterme b und a berücksichtigen die aus der verschiedenen Ellipsenexzentrizität herrührenden verschiedenen Energien der Anfangs- und Endzustände n bzw. n'. Anschaulich entstehen die Energieunterschiede durch die verschieden starke Wechselwirkung des Leuchtelektrons mit dem Atomrumpf. Mit zunehmender Exzentrizität der Ellipsenbahn dringen die Elektronen in den Atomrumpf ein, in Gebiete schwächerer Kernabschirmung.

Erlaubte, verbotene Übergänge. Nicht alle Kombinationen zwischen den Termen der Alkalien nach Abb. 5.6.1–4 treten als Spektrallinien in Erscheinung. Man beobachtet wesentlich weniger Linien und unterscheidet so zwischen erlaubten und verbotenen Übergängen. Die Definition besagt, daß Spektrallinien, die nicht beobachtet werden, bzw. nur mit einer um Größenordnungen geringeren Intensität auftreten, zu verbotenen Übergängen gehören. Man findet, daß die Drehimpulsquantenzahl für erlaubte Übergänge sich nur um eine Einheit ändert ($\Delta l = \pm 1$). Verboten sind Übergänge zwischen gleichen Drehimpulsquantenzahlen und für Änderungen um mehr als eine Einheit ($\Delta l = 0$, $|\Delta l| > 1$). Ursache ist der Drehimpulserhaltungssatz. Übergänge sind mit der Emission bzw. Absorption eines Lichtquantes verbunden. Das Lichtquant hat den Drehimpulswert $1\hbar$. Dieser Drehimpuls wird entweder dem Atom abgeführt oder zugeführt. Deshalb gelten diese Auswahlregeln auch nicht z. B. für Anregungen mittels Elektronenstoß. Die Übergänge sind nur „optisch" erlaubt bzw. verboten.

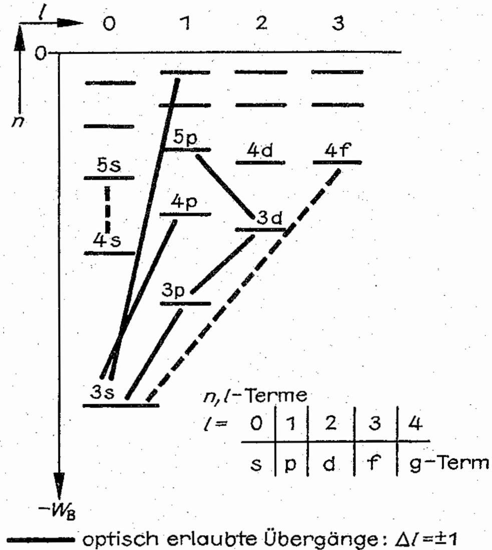


Abb. 5.6.1–4 Termschema der Alkaliatome am Beispiel des Natriums

5.6.1.6. Die Orientierungsquantenzahl. Ursprünglich gibt es keinen Grund, Aussagen zur Orientierung der Ellipsenbahnen im Raum zu machen. Das ändert sich, wenn die Atome einem Magnetfeld ausgesetzt werden. Die im Atom kreisenden Elektronen bilden ein Magnetfeld aus. Der Effekt ist analog dem bei einer stromdurchflossenen Spule. Unter der Wirkung des äußeren Magnetfeldes richtet sich das atomare Magnetfeld aus. Die Kreise bzw. die Ellipsen sind dadurch im Raum orientiert. Wieder ist eine Quantenbedingung zu beachten. Die Komponente des Drehimpulses l_Z in Richtung des äußeren Magnetfeldes kann nur bestimmte Werte annehmen, $|l_Z| = m\hbar$; $|\pm m| \leq l < \sqrt{l(l+1)}$, $m = 0, -1, +1, -2, +2, \ldots, -l, +l$ (m = Magnetquantenzahl oder auch Orientierungsquantenzahl). m umfaßt $2l+1$ mögliche Werte. Die größten Werte gehören zur Ausrichtung des Drehimpulses in Richtung des äußeren Magnetfeldes, bzw. in die entgegengesetzte Richtung. Der Wert $m = 0$ entspricht einem Drehimpuls senkrecht zum äußeren Magnetfeld.

Das magnetische Moment des Atoms ist ein Maß für die Änderung der Bahnenergie der Elektronen unter der Wirkung eines äußeren Feldes. Die Änderung der Bahnenergie bewirkt das Auftreten verschiedener Spektrallinien zu einem Drehimpulswert, die Spektrallinien spalten im Magnetfeld in $2l+1$ Linien auf. Das *magnetische Bahnmoment* μ_l eines Elektrons ergibt sich zu $\mu_l = -\sqrt{l(l+1)} \cdot e\hbar/2m_e$ (m_e = Elektronenmasse, l = Bahndrehimpulsquantenzahl). Die Grundgröße $e\hbar/2m_e = \mu_B$ ist das *Bohrsche Magneton*, $\mu_B = 9{,}274 \cdot 10^{-24}\,\mathrm{Am^2}$.

Die Energiedifferenz der Elektronenbahnen beträgt bei gleichem Bahndrehimpuls $\Delta E = m\mu_B H$ (m = Magnetquantenzahl, H = Feldstärke des äußeren Magnetfeldes). Diese Aufspaltung des Spektrallinien bezeichnet man als *Zeeman-Effekt*. Die Aufspaltung der Spektrallinien in elektrischen Feldern vollzieht sich analog und trägt den Namen *Stark-Effekt*.

5.6.2. Linienfein- und Hyperfeinstrukturen

5.6.2.1. Der Spin des Elektrons. Nach den bisher dargestellten Befunden aus der Analyse der Spektrallinien wird die Energie eines Elektronenzustandes $W_{n,l,m}$ durch die drei Quantenzahlen n, l und m festgelegt. Dabei ist noch nicht erwähnt, daß die Spektrallinien des Wasserstoffes bei genauerem Hinsehen fast alle aus zwei eng benachbarten Linien bestehen. Auch die besprochenen Spektren der Alkalien haben noch diese Feinstruktur. Bekannt ist die Aufspaltung der gelben Na-D-Linie. Der Wellenlängenunterschied beträgt $\approx 0{,}6$ nm, das entspricht einer Energiedifferenz von $\Delta W = 0{,}0021$ eV. Dieser Sachverhalt und viele andere experimentelle Tatsachen führten zum Auffinden einer der grundlegenden Elementarteilcheneigenschaften. Das Elektron hat einen Eigendrehimpuls, *Spin* genannt. Auch die Bezeichnung *Spindrehimpuls* ist üblich. Die Planetenmodellvorstellungen weitergeführt, dreht sich das Elektron noch um seine eigene Achse, analog zur Erdrotation. Die Analogiebetrachtung hat aber enge Grenzen, wie später noch gezeigt werden wird. Der Spin des Elektrons ist nur eines Wertes fähig. Er hat einen festen Betrag, der nicht veränderlich ist, eine echte Elementarteilcheneigenschaft darstellt. Die Größe des Spins ist nur mittels der Quantentheorie richtig berechenbar. Eigendrehung und Ladung führen zu einem magnetischen Moment. Das magnetische Moment des Elektrons koppelt mit seinem magnetischen Bahnmoment. Je nachdem, wie die Momente koppeln, entstehen verschiedene Energiezustände der Elektronen im Atom, woraus die Feinstruktur der Spektrallinien resultiert. Der Spindrehimpuls s unterliegt der allgemeinen quantentheoretischen Gesetzmäßigkeit durch die Form $|s| = \sqrt{s(s+1)}\,\hbar$. Die Spinquantenzahl s ist aber nur eines Wertes fähig, $s = 1/2$. Die zugehörigen Einstellmöglichkeiten im Magnetfeld sind durch die Z-Komponenten des Eigendrehimpulses s_Z vom Betrage $1/2\hbar$ gegeben. Der Eigendrehimpuls kann sich nur parallel oder antiparallel zum Bahndrehimpuls einstellen. Der Gesamtdrehimpuls $\boldsymbol{j}$ ist die Vektorsumme von Spin- und Bahndrehimpuls, $\boldsymbol{j} = \boldsymbol{s} + \boldsymbol{l}$. Die Quantenzahl des Gesamtdrehimpulses ist demnach $j = l \pm 1/2$.

Das bedeutet, daß alle Terme der bisher betrachteten Elektronenzustände im Atom zwei Zuständen entsprechen, einem, der den Gesamtdrehimpuls, bezogen auf den Bahndrehimpuls, vergrößert, und einem, der ihn kleiner werden läßt. Auch die Bezeichnung „Spin auf", „Spin ab" ist üblich. Die Bezeichnung, eine auf den Bahndrehimpuls bezogene Spineinstellung „parallel" bzw. „antiparallel" vorliegen zu haben, ist nicht exakt. Existiert kein Bahndrehimpuls ($l = 0$), so hat der Spindrehimpuls keine Einstellmöglichkeiten. Deshalb spalten die $l = 0$-Terme (s-Zustände) nicht auf, der Zustand bleibt entartet. Die Entartung kann in einem äußeren Magnetfeld aufgehoben werden.

Das Termschema nach Abb. 5.6.1–4 ist unter Berücksichtigung der Feinstrukturaufspaltung gemäß Abb. 5.6.2–1 zu erweitern. Der Gesamtdrehimpuls wird in der Termbezeichnung als Index unten angefügt. $3p_{3/2}$ bzw. $3p_{1/2}$ sind die Bezeichnungen von zwei Energiezuständen mit $n = 3$, $l = 1$ und $j = l + s = 3/2$ bzw. $j = l - s = 1/2$.

Die Auswahlregeln für optisch erlaubte Übergänge ($\Delta l = \pm 1$) ergeben für die Änderung der Quantenzahl des Gesamtdrehimpulses die Werte $\Delta j = 0 \pm 1$. Dabei ent-

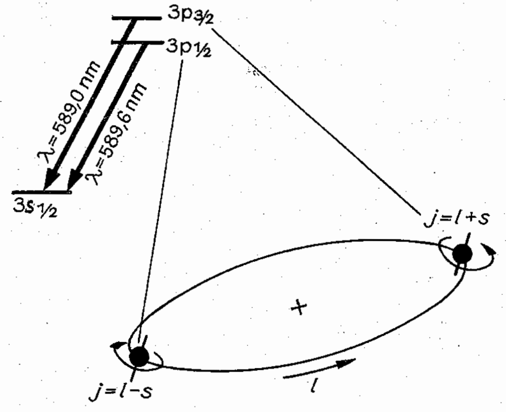


Abb. 5.6.2–1 Spin-Bahn-Kopplung und Feinstrukturaufspaltung der dicken Natrium-Deuterium-Linie

spricht $\Delta j = 0$ einer Umkehr der Spinrichtung ohne Änderung des Bahndrehimpulses. Man spricht von einem „Umklappen" des Spins.
Das aus dem Spin des Elektrons resultierende magnetische Moment des Elektron μ_s ergibt sich quantentheoretisch zu $\mu_s = -\sqrt{s(s+1)}\, e\hbar/m_e$. Der Wert ist doppelt so groß, wie er nach den Vorstellungen der klassischen Physik für die Eigenrotation eines Teilchens sein müßte.

5.6.2.2. Der Nachweis des Elektronenspins. 1926 fügten Goudsmid und Uhlenbeck die Spinquantenzahl den drei bis dahin existierenden Quantenzahlen hinzu und äußerten die Vermutung, daß das Elektron einen inneren Spindrehimpuls hat. Einen ersten Hinweis auf die Existenz des Eigendrehimpulses kann man schon den Versuchen von Stern und Gerlach zur Aufspaltung von Atomstrahlen im inhomogenen Magnetfeld (1921) entnehmen. Die magnetischen Momente der Atome mit abgeschlossenen Elektronenschalen kompensieren sich zu Null. Silberatome haben derart abgeschlossene Schalen und je ein Leuchtelektron im $5s_{1/2}$-Zustand. Das eine Leuchtelektron je Atom hat aufgrund seines Drehimpulswertes $l = 0$ kein magnetisches Bahnmoment. Kann ein magnetisches Moment nachgewiesen werden, so ist es nur auf einen Eigendrehimpuls der Elektronen zurückzuführen.
Stern und Gerlach erzeugten einen Strahl von Silberatomen und ließen ihn durch ein sehr starkes, sehr inhomogenes Magnetfeld dringen. Danach traf der Silberdampfstrahl auf eine Glasplatte, wo der Niederschlag direkt sichtbar war. Abb. 5.6.2–2 skizziert den natürlich im Vakuum ablaufenden Versuch. Bei eingeschaltetem Magnetfeld wird eine Ablenkung des Strahls beobachtet. Der Strahl spaltet in zwei Teilstrahlen auf. Es werden weder verschiedenstarke Ablenkungen noch Ablenkungen in mehr als diese zwei Richtungen festgestellt, und alle Atome unterliegen der Ablenkung.
Im inhomogenen Magnetfeld greift an einem magnetischen Dipol mit dem Moment μ eine Kraft K_z an, die proportional μ_z und der Feldänderung $\mathrm{d}\boldsymbol{H}/\mathrm{d}z$ ist. Die Elektronen der Silberatome richten sich im Magnetfeld aus. Sie stellen sich mit ihrem magnetischen Moment parallel und antiparallel zum äußeren Magnetfeld ein. Je nach ihrer Einstellung werden sie in Richtung z oder $-z$ abgelenkt. Die Strahlenaufspaltung ist scharf, d. h. es existiert nur ein Betrag für das magnetische Moment, womit auch der Betrag des Spindrehimpulses nur einen Wert hat. Da alle Atome abgelenkt werden, kann der Spin nur zwei Einstellmöglichkeiten haben. Sie entsprechen der Spinquantenzahl $s = +1/2$ und $-1/2$. Damit ergeben sich die nur zwei möglichen Werte für die Komponenten des magnetischen Moments, $\mu_Z = s_Z e\hbar/m = \pm\mu_B$. Der Versuch ist mit vielen Elementen nachvollzogen worden. Durch den Elektronenspin ist auch verständlich, warum atomarer Wasserstoff paramagnetisch ist. Paramagnetisch sind Materialien, die ihre atomaren Magnetfelder in einem äußeren Magnetfeld ausrichten. Sie verhalten sich ähnlich wie Dielektrika mit Orientierungspolarisation. Der Grundzustand der Wasserstoffelektronen ist ein s-Zustand ($l = 0$), sie haben keinen Bahndrehimpuls und kein magnetisches Bahnmoment. Das atomare Magnetfeld kann nur durch den Eigendrehimpuls der Elektronen gebildet sein.

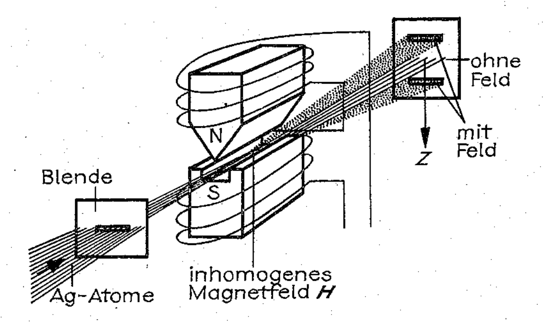


Abb. 5.6.2–2 Stern-Gerlach-Versuch; Nachweis der Elektronen durch Ablenkung von Silberatomen im inhomogenen Magnetfeld

5.6.2.3. Hyperfeinstruktureffekte. Die Atomspektroskopie mit immer höherer Auflösung betreibend, gelangte man zu Effekten, die die durch die Feinstrukturaufspaltung der Spektrallinien entstandenen Linien wiederum aufspalten ließen. Zur Erklärung dieser *Hyperfeinstruktur* reichen die vier Quantenzahlen n, l, m, s der Elektronenzustände im Atom nicht aus. Eine der Ursachen liegt in der Existenz eines Eigendrehimpulses der Atomkerne. Die Elementarteilchen des Kernes, Protonen und Neutronen, haben den Drehimpuls $i = 1/2\hbar$. Bei allen Kernen mit ungerader und einigen mit gerader Nukleonenzahl kompensieren sich die Kernspins nicht zu 0. Die daraus resultierenden magnetischen Momente koppeln mit den Spinmomenten der Elektronen. Das daraus entstehende gesamte magnetische Spinmoment des Atoms bildet sich wie das für die Feinstrukturaufspaltung betrachtete gesamtmagnetische Moment der Elektronen. Auch die Einstellmöglichkeiten in einem äußeren Magnetfeld sind analog. Das kernmagnetische Moment ist um drei Größenordnungen geringer als das magnetische Moment der Elektronen. Entsprechend klein ist die Wechselwirkung zwischen Kern und Hülle. Die dadurch bedingte Aufspaltung der Spektrallinien beträgt im Energiemaßstab $\sim 10^{-5}$ eV.
Die in Abb. 5.6.2–1 dargestellte Feinstrukturaufspaltung der gelben Na-D-Linie in die Linien D_1 und D_2 liefert noch nicht die endgültige Linienzahl. Beide Linien zeigen eine Hyperfeinstruktur mit Wellenlängendifferenzen von $\Delta\lambda = 0{,}0023$ bzw. $0{,}0021$ nm. Experimente zur Hyperfeinstrukturaufspaltung gestatten die Bestimmung der kernmagnetischen Momente und des Spinzustandes der Kerne. Der Kern des Na-Atoms hat im Grundzustand den Spin $3/2\hbar$.
Neben dem Kernspin sind noch andere Kerneigenschaften für hyperfeinstrukturartige Aufspaltungen der Spektrallinien verantwortlich. In vielen Fällen ist die Ladungsverteilung im Kern nicht mehr kugelsymmetrisch, auch ist sie unterschiedlich bei verschiedener Neutronenzahl. Die dem Kern besonders nahe kommenden s-Elektronen „spüren" diese verschiedene Ladungsverteilung, und ihre Energieniveaus sind für verschiedene Iostope verschieden. Ein anderer *Isotopieeffekt* entsteht durch die Massenabhängigkeit der Rydberg-Konstanten in der Formel für die Frequenz der Spektrallinien. Die Massenabhängigkeit hat ihre Ursache in einer leichten Kernmitbewegung, hervorgerufen durch die kreisenden Elektronen (vgl. 5.6.1.3.). Der Effekt ist um so größer, je leichter die Kerne sind. Der Unterschied in den Spektrallinien der Wasserstoffisotope führte zum erstmaligen Nachweis des Deuteriums durch Urey im Jahre 1932.

5.6.2.4. Mehrelektronensysteme. In allem Bisherigen war immer nur ein äußeres Elektron betrachtet. Außerhalb des Atomrumpfes können auch zwei und mehr Elektronen an-

geordnet sein. Gerechtfertigt bleibt die Annahme, daß abgeschlossene Elektronenschalen existieren. In ihnen kompensieren sich der Drehimpuls und das magnetische Moment zu Null, und die Kernladung ist außerhalb dieses Rumpfes um die Elektronenzahl dort reduziert. Auch Elektronen aus abgeschlossenen Schalen können in angeregte Zustände gehoben werden, was aber größere Energiebeträge erfordert, wodurch diese Spektren nicht mit denen der äußeren Elektronen überlappen.

Je mehr Elektronen außerhalb des Atomrumpfes existieren, um so komplizierter werden die Spektren. Mehrere Termsysteme mit auch großer Linienvielfalt treten auf. Beispielsweise beim Helium mit seinen zwei Elektronen existiert das Termsystem des Ortho- und das des Parheliums. Beim Parhelium sind die Spins der beiden Elektronen einander entgegengerichtet und beim Orthohelium gleichgerichtet.

Die Spin- und Bahndrehimpulse aller äußeren Elektronen wechselwirken über ihre magnetischen Momente miteinander. Zwei Kopplungsarten treten auf, man kann sie als Grenzfälle einer Vielfalt der Einstellmögichkeiten betrachten. Es sind dies die *j-j-Kopplung* und die *L-S-Kopplung*. Wenn bei jedem äußeren Elektron erst einmal Spindrehimpuls s und Bahndrehimpuls l den Gesamtdrehimpuls j, $j = l + s$ bilden, so äußert sich die Wechselwirkung der Elektronen miteinander in der vektoriellen Addition ihrer Gesamtdrehimpulse. Der Gesamtdrehimpuls der Hülle entsteht durch $J = \sum_i j_i$, $j_i = l_i + s_i$ (*j-j-Kopplung*).

Meistens realisiert sich eine andere Form. Es koppeln erst alle Spins s_i und alle Bahndrehimpulse l_i untereinander zum Gesamtspin S der Hülle bzw. zum Gesamtbahndrehimpuls L der Hülle und erst diese dann miteinander, $J = L + S$, $L = \sum_i l_i$, $S = \sum_i s_i$ (*L-S-Kopplung*). Für diese Art der Bildung des Gesamtdrehimpulses J der Hülle ist auch die Bezeichnung Russell-Saunders-Kopplung gebräuchlich.

Das resultierende magnetische Moment der äußeren Elektronen bildet sich mit der Gesamtdrehimpulsquantenzahl J zu $M_J = g(L, S, J) \cdot \sqrt{J(J+1)}\, \mu_B$. Die Einheit ist wieder das Bohrsche Magneton μ_B. In der üblichen, durch die Quantentheorie begründete Art, ist die Gesamtdrehimpulsquantenzahl J für das magnetische Moment M_J in Richtung J einzusetzen. g heißt *Landé-Faktor*. Er ist aus den Quantenzahlen L, S und J berechenbar und entsteht durch die Vektoraddition von L und S zu J. Für den Sonderfall des reinen Spinmagnetismus ($L = 0$) beträgt $g = 2{,}0023$, für reinen Bahnmagnetismus ($S = 0$) gilt $g = 1$. Der Landé-Faktor g ist für reinen Spinmagnetismus zum einen etwa doppelt so groß wie für reinen Bahnmagnetismus (bei gleichgroßen Quantenzahlen), zum anderen nicht genau ganzzahlig. Die Erklärung dafür liefert erst der Formalismus einer quantentheoretischen Behandlung.

5.6.2.5. Breite der Spektrallinien. Die Frequenz einer Spektrallinie kann sehr genau gemessen werden. Zur Messung der Linienaufspaltung im Magnetfeld (Zeeman-Effekt) ist es beispielsweise notwendig, Linien mit Wellenlängenunterschieden von $\Delta\lambda \sim 10^{-9}$ m zu trennen. Bei der Hyperfeinstrukturaufspaltung unterscheidet man Linien im Abstand von $\Delta\lambda \sim 10^{-12}$ m. So ergibt sich die Frage, ob eine Grenze für das Auflösungsvermögen existiert, oder ob mit immer besseren Geräten grenzenlos immer genauer gemessen werden kann, ob eventuell immer wieder immer feinere Strukturen und dabei immer neue Effekte entdeckt werden können. Es zeigt sich, daß die Spektrallinien eine gewisse Breite, eine Frequenzausdehnung aufweisen und nicht unendlich scharf sind. Dadurch überlappen dicht beieinander liegende Linien und sind bei zu geringem Frequenzunterschied nicht mehr in einzelne Linien aufzulösen. Abb. 5.6.2–3 dient der Erklärung des Begriffes *Halbwertsbreite*, einer Meßgröße zur Beschreibung der Linienbreite. Die Halbwertsbreite $\Delta\nu_{HWB}$ ist die Ausdehnung einer Verteilung in Höhe des halben Maximalwertes.

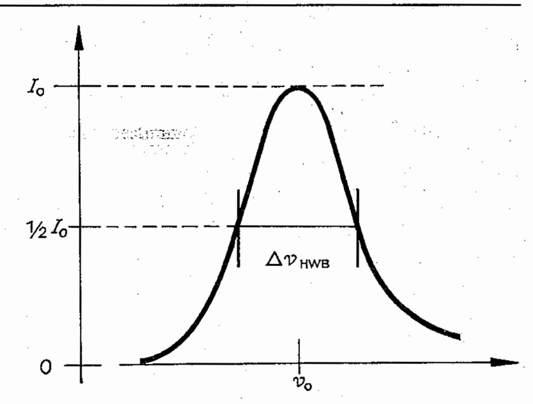


Abb. 5.6.2–3 Halbwertsbreite einer Verteilung

Die Art der Lichtentstehung läßt schon vermuten, daß Spektrallinien nicht unendlich schmal sein können. Licht besteht aus einzelnen Quanten (vgl. 5.5.5.), die in unregelmäßiger Aufeinanderfolge (statistisch verteilt) von Atomen abgestrahlt werden. Die Wellenzüge schließen nicht aneinander an, es entsteht kein unendlich langer Wellenzug, kein kohärentes Licht. Endliche Wellenzüge sind aber nur durch eine Verteilung mindestens mehrerer Frequenzen beschreibbar (vgl. 2.6.3.).

Quantentheoretisch liefert die Heisenbergsche Unbestimmtheitsbeziehung (vgl. 5.5.4.4.) eine von der Meßmethode unabhängige Unschärfe für die abgestrahlte Energie. Es gilt $\Delta E \cdot \tau \approx \hbar$; die Schwankungsbreite der Energie (ΔE) kann als Produkt mit der Lebensdauer τ der das Licht abstrahlenden angeregten Zustände nie geringer als etwa $\hbar$ sein ($\hbar = h/2\pi$; h = Plancksches Wirkungsquantum). Normale angeregte Zustände kehren nach $\approx 10^{-8}$ s in den Grundzustand zurück. Demnach existiert für sie eine „natürliche Energiebreite" ΔW von $\approx 10^{-7}$ eV ($h = 4{,}13 \cdot 10^{-15}$ eV · s). Mit $\Delta W = h\Delta\nu$ ergibt sich aus der Heisenbergschen Unbestimmtheitsrelation: $\Delta\nu = 1/(2\pi\tau)$ (*natürliche Linienbreite*) eine naturgemäß existierende, nicht zu verringernde Linienbreite der Spektrallinien. Sie ist durch die mittlere Lebensdauer des die Strahlung aussendenden Niveaus bestimmt. Es existieren auch langlebige, strahlende Energieniveaus, zu denen äußerst geringe natürliche Linienbreiten gehören (vgl. 5.2.4.4.).

Die natürliche Linienbreite ist sehr gering, aber meßbar, oft wird sie von anderen, die Linie verbreiternden Effekten überdeckt.

Doppler-Verbreiterung. Die das Licht emittierenden Atome befinden sich normalerweise nicht in Ruhe. Sie führen entsprechend ihrer thermischen Energie ungeordnete Bewegungen aus, besonders ausgeprägt bei Gasen. Wegen des Doppler-Effektes (vgl. 3.1.1.3.) verändert sich die Frequenz in Abhängigkeit von der Geschwindigkeitskomponente des Atoms in Emissionsrichtung. Für ein strahlendes Gas führt die Doppler-Verbreiterung bei Zimmertemperatur zu Halbwertsbreiten, die um drei Größenordnungen

über der natürlichen Linienbreite liegen. Die Doppler-Verbreiterung ist von anderen Effekten durch ihre Proportionalität mit $A^{-1/2}$ (A = Massenzahl der Atome) zu unterscheiden.

Stoßverbreiterung führt zu ähnlich drastischer Linienaufweitung. Die sich regellos bewegenden Atome nicht allzustark verdünnter Gase stoßen miteinander. Auch während des Emissionsvorganges sind mehrere Stöße wahrscheinlich. Das Licht erleidet dabei Phasensprünge, und auch die Emissionsrichtung wird verändert. Typische stoßverbreiterte Linien entstehen z. B. in Quecksilberhöchstdrucklampen.

5.6.3. Das Periodensystem der Elemente (PSE)

Systematisierung gehört zu den ersten Schritten eines wissenschaftlichen Herangehens an Datenmengen.

Im Altertum schon ordnete Aristoteles, Gegensätze betonend, die damaligen Elemente und Grundeigenschaften (Abb. 5.6.3–1a). Zu Beginn des 19. Jahrhunderts waren für genügend Elemente die chemischen Eigenschaften und

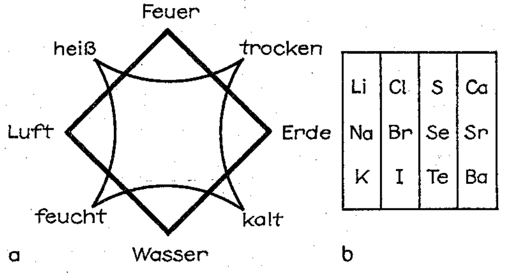


Abb. 5.6.3–1 (a) Ordnung der Elemente und Grundeigenschaften nach Aristoteles, (b) Anordnung chemischer Elemente in Dreiergruppen durch Döbereiner

Massenzahlen bekannt, um damit eine Grundlage für die heute gebräuchliche Darstellung aller Elemente bilden zu können. Döbereiner systemmatisierte Elemente nach ihren ähnlichen chemischen Eigenschaften in Dreiergruppen, wobei die Atommassendifferenzen innerhalb der Dreiergruppen annähernd gleich sind (Abb. 5.6.3–1b). Die heutige gebräuchlichste Form des *Periodensystems der Elemente* wurde durch Meyer und Mendelejew unabhängig voneinander formuliert. In ihm sind die Elemente zum einen nach steigender Protonenzahl, zum anderen nach gleichen chemischen Eigenschaften aufgetragen. Nach rechts steigt die Protonenzahl (Ordnungszahl). Die periodische Wiederkehr von Elementen mit ähnlichen chemischen Eigenschaften wird durch Beginn einer neuen Zeile so berücksichtigt, daß diese untereinander zu finden sind. Eine atomtheoretische Begründung solch einer Darstellung ist möglich. Sie ergibt sich aus den Gesetzmäßigkeiten des Aufbaus der Elektronenhülle der Atome. Stehen nicht die chemischen Eigenschaften der Atome im Vordergrund, so sind mitunter andere Systematisierungen praktisch. In der Kernphysik sind die chemischen Eigenschaften fast bedeutungslos, die dort bevorzugte Darstellung der Elemente und Isotope (Nuklidkarte) läßt alle kernphysikalischen Eigenschaften übersichtlich erscheinen. Für andere Fälle sind Tabellen nützlich, etwa zur Auflistung der relativen Atommassen oder der Elektronenkonfigurationen.

5.6.3.1. Die vier Quantenzahlen. Das Bohrsche Atommodell führt zu vier Quantenzahlen für die Beschreibung aller möglichen Elektronenzustände im Atom (vgl. 5.6.1.). Im Modellfall des Atoms mit einem Elektron sind diese Elektronenzustände mögliche Anregungsstufen. Das eine Elektron kann durch Energiezufuhr in jeden der Anregungszustände gehoben werden und dort strahlungsfrei umlaufen. Ihm ist das Bestreben eigen, ohne äußere Einwirkung (spontan), unter Energieabstrahlung in den tiefstmöglichen Anregungszustand, in den Grundzustand, zurückzukehren.

Die durch die vier Quantenzahlen dargestellten möglichen Elektronenbahnen haben eine noch weitaus tiefgründigere Bedeutung. Belegt man jede Bahn, bei der kernnähesten angefangen, mit je einem Elektron und ergänzt dabei den Kern immer um jeweils ein Proton und auch auf die notwendige Neutronenzahl, so bilden sich nach und nach alle chemischen Elemente. Also: Die Zustände sind nicht nur Anregungszustände für ein Elektron, sie können aufgefüllt werden mit je einem Elektron, wodurch alle möglichen Atome entstehen, sich die Periodizitäten im PSE deuten lassen und chemische Eigenschaften aus dem Aufbau der Elektronenhülle erklärbar werden.

Hauptquantenzahl n kennzeichnet den mittleren Abstand des Elektrons vom Kern. Mit ihr ist grob die Energie gegeben, da alle anderen Quantenzahlen diese nicht so stark beeinflussen. Damit sind alle Niveaus mit gleicher Hauptquantenzahl n energetisch benachbart und bilden eine „Schale" (Schalenmodell des Atomaufbaues). Die Schalenbezeichnung K, L, M, N, O, P, Q zu den Hauptquantenzahlen $n = 1, 2, 3, 4, 5, 6, 7$ bildet gleichzeitig die Bezeichnung der Perioden (Zeilen) im PSE, was im weiteren noch verständlich werden wird. Hauptquantenzahl und Zeilennummer im PSE stimmen überein.

Bahndrehimpulsquantenzahl l führt zur Unterteilung der Schalen in Unterschalen. l wird auch *Nebenquantenzahl* genannt. In der anschaulichen Deutung beschreibt l die zu einem Wert von n möglichen Ellipsenbahnen. Sie unterscheiden sich auch in der Energie. Verschiedene l-Werte kennzeichnen verschiedene kleine Halbachsen bzw. Exzentrizitäten. l durchläuft alle Werte von 0 bis $l - 1$. Für $l = 0$, 1, 2, 3 sind die Bezeichnungen s-, p-, d-, f-Term bzw. -Elektron üblich (vgl. 5.6.1.5.). Der Wertevorrat, die Anzahl der Unterschalen ist n.

Magnetquantenzahl m gibt die mögliche Zahl der Ellipsen gleicher Form und Größe (gleich n- und l-Werte) an. Sie entsteht durch die verschiedenen Möglichkeiten der Orientierung im Raum, welche durch die magnetische Wechselwirkung erwirkt wird. m wird auch als Orientierungsquantenzahl bezeichnet. Sie durchläuft alle Werte von l über 0 bis -1; $m = 0, \pm 1, \pm 2, \ldots, \pm l - 1, \pm l$. Der Wertevorrat beträgt $2l + 1$. Im Termsystem bedeutet das, daß jede Unterschale noch einmal feinstrukturiert ist, sie enthält $2l + 1$ Terme.

Spinquantenzahl s kann nur zwei Werte annehmen, $s = +1/2$ und $-1/2$. Sie kennzeichnet die zwei möglichen Richtungen des Eigendrehimpulses der Elektronen relativ zum Bahndrehimpuls. Mit ihr verdoppelt sich die Anzahl der Anregungs- bzw. Besetzungsmöglichkeiten aller bisher angeführten Zustände. Damit enthält jede Unterschale $2(2l + 1)$ Terme. s- und p-Terme addiert ($l = 0$ und 1), ergibt die Zahl der Gruppen (Spalten) im PSE. In der Summation aller Unterschalen beträgt für eine durch n gekennzeichnete Schale die Anzahl der Terme $\sum_{l=0}^{n-1} 2(2l + 1) = 2n^2$.

Abb. 5.6.3–2 stellt die möglichen Elektronenzustände dar, gibt die Bedeutung der Quantenzahlen im Modell der Elektronenbahnen nach BOHR und SOMMERFELD an und zeigt in der letzten Spalte das Element an, welches bei Besetzung aller niederen Terme mit Elektronen entsteht.

5.6.3.2. Allgemeines Stabilitätsprinzip und Pauli-Prinzip. Für das physikalische Aufbauschema der Atome sind neben den vier Quantenzahlen noch zwei Bedingungen wesentlich:

Allgemeines Stabilitätsprinzip. Es besagt, daß die tiefstliegenden Energieniveaus zuerst besetzt werden. Die energetische Reihenfolge der Terme ist dementsprechend aus den Elektronenkonfigurationen nach Tab. 5.6.3–3 ersichtlich. Die geordnete Reihenfolge nach Abb. 5.6.3–2 zur Besetzung der Terme endet beim Argon. Für das nachfolgende Element Kalium wird das letzte Elektron nicht in die 3d-Schale eingebaut. Die 4s-Terme liegen energetisch tiefer, der erste Term wird für das Kalium besetzt, der zweite für das Element Calcium. Erst dann wird mit Scandium beginnend, die 3d-Schale aufgefüllt. Zu schwereren Elementen hin wird öfters mit dem Aufbau neuer Unterschalen begonnen, ehe die vorhergehenden aufgefüllt sind. Die Reihenfolge ergibt sich nur aus der energetischen Lage der Elektronenniveaus. Das ist zum einen mit großem Rechenaufwand quantentheoretisch berechenbar und für viele Fälle schon berechnet, zum anderen durch Messung der Ionisierungsenergien bewiesen.

Pauli-Prinzip. Es formuliert den Sachverhalt, daß jeder durch die vier Quantenzahlen gegebene Zustand nur von einem Elektron besetzt werden kann. Man kann auch sagen: In einem Atom existieren nur Elektronen, die sich in mindestens einer Quantenzahl unterscheiden. Das Pauli-Prinzip ist wesentlich allgemeiner faßbar, es gilt auch für andere Teilchen mit halbzahligem Spin und ist auch ein Aufbauprinzip des Atomkerns, der aus den Spin-1/2-Teilchen Protonen und Neutronen besteht. Die Quantentheorie ordnet ihnen antisymmetrische Zustandsfunktionen zu und bildet mit ihnen eine als Fermionen bezeichnete Teilchenklasse. Andere Elementarteilchen haben ganzzahligen Spin und gehören mit ihren symmetrischen Wellenfunktionen zur Klasse der Bosonen. Für sie gilt das Pauli-Prinzip nicht.

5.6.3.3. Erklärung der Ordnung und Periodizität im PSE. Der von MEYER und MENDELEJEW gefundene systematische Aufbau der chemischen Elemente erklärt sich durch die Anordnung der äußeren Elektronen. Elektronen innerer Schalen sind so fest gebunden, daß sie nicht auf die dagegen geringen Energien chemischer Reaktionen (Änderungen der Bindung) reagieren. Dem entsprechend ist ein Element mit abgeschlossener Schale besonders stabil. Schalenabschlüsse sind bei den Edelgasen Helium, Neon, Argon, Krypton, Xenon und Radon erreicht, hier sind gerade alle s- und p-Terme besetzt. Eine neue Schale beginnt immer mit der Besetzung eines s-Zustandes. Die Elemente mit einem s-Elektron Wasserstoff, Lithium, Natrium, Kalium, Rubidium, Caesium und Francium stehen in der ersten Spalte des PSE, sie werden als Alkaliatome bezeichnet. Ihr einziges äußeres Elektron ist noch relativ locker gebunden, woraus die besonders gute chemische Reaktionsfähigkeit resultiert. Sind beide s-Zustände mit Elektronen besetzt, so ist immer die zweite Spalte im PSE erreicht. Ihre Elemente bilden die Gruppe der Erdalkalien. Bei ihnen beginnt, wie in Tab. 5.6.3–3 gestrichelt gekennzeichnet, die Bildung der Nebengruppenelemente. Das sind Elemente mit aufgefüllten 2s-Termen, bei denen die Besetzung der d-Terme nachgeholt wird. Bei den Elementen der P- und Q-Schale ist die Besetzung der d-Terme noch einmal nach dem ersten Element der Nebengruppe Lanthan bzw. Actinium unterbrochen. Danach werden erst die Elemente mit Elektronen auf f-Termen gebildet, die Lanthaniden bzw. Actiniden (in der Tabelle gepunktet). Sie sind untereinander chemisch sehr ähnlich, da sie alle zwei s-Elektronen in der äußeren Schale haben und auch kein großer Unterschied in der Besetzung der d-Terme besteht.

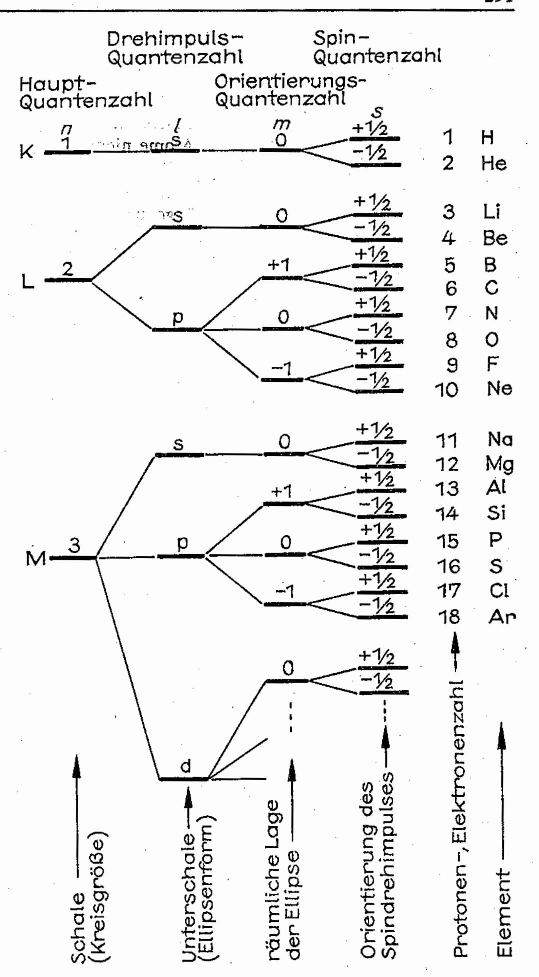


Abb. 5.6.3–2 Termdarstellung der möglichen Elektronenzustände, ihre Bedeutung im Bohrschen Atommodell und Elektronenkonfigurationen der Atome

Die genaue Betrachtung dieses Schalenaufbaus führt tief in das Verständnis der chemischen Eigenschaften der Elemente. Hierfür sei nur ein Beispiel angeführt. Die Wertigkeit der Elemente ist durch die Zahl der Elektronen in äußeren s- und p-Zuständen gegeben. Kupfer ist nach der Tabelle einwertig (ein 4s-Elektron). Chemische Bindungen können die Terme in ihrer energetischen Lage ein wenig verändern. Im Falle des Kupfers wird es für das zehnte d-Elektron energetisch günstiger, in den zweiten 4s-Zustand überzugehen, wodurch Cu dann zweiwertig wird.

Viele physikalische und chemische Eigenschaften spiegeln die durch den Schalenaufbau begründete Periodizität der Elemente wider. Will man ein Elektron eines neutralen Atoms entfernen, so ist dazu die *Ionisationsenergie* aufzubringen. Sie ist ein Maß für die Stärke der Bindung des äußersten Elektrons. In Abb. 5.6.3–4 ist die Ionisationsenergie $W_{\text{Ionisation}}$ der Elemente des Periodensystems bis Caesium aufgetragen. Die Stabilität der Edelgase zeigt sich in ihren auffallend hohen Ionisationsenergien. Besonders leicht abzutrennen ist das einzelne äußere Elektron der Alkaliatome (Lithium, Natrium, Rubidium, Caesium). Auch feinere Details des Schalenaufbaues sind dem Bild zu entnehmen.

Tab. 5.6.3–3 Elektronenkonfigurationen der Elemente

		K	L		M			N				O	
		1s	2s	2p	3s	3p	3d	4s	4p	4d	4f	5s	5p
1	H	1											
2	He	2											
3	Li	2	1										
4	Be	2	2										
5	B	2	2	1									
6	C	2	2	2									
7	N	2	2	3									
8	O	2	2	4									
9	F	2	2	5									
10	Ne	2	2	6									
11	Na	2	2	6	1								
12	Mg	2	2	6	2								
13	Al	2	2	6	2	1							
14	Si	2	2	6	2	2							
15	P	2	2	6	2	3							
16	S	2	2	6	2	4							
17	Cl	2	2	6	2	5							
18	Ar	2	2	6	2	6							
19	K	2	2	6	2	6		1					
20	Ca	2	2	6	2	6		2					
21	Sc	2	2	6	2	6	1	2					
22	Ti	2	2	6	2	6	2	2					
23	V	2	2	6	2	6	3	2					
24	Cr	2	2	6	2	6	4	2					
25	Mn	2	2	6	2	6	5	2					
26	Fe	2	2	6	2	6	6	2					
27	Co	2	2	6	2	6	7	2					
28	Ni	2	2	6	2	6	8	2					
29	Cu	2	2	6	2	6	10	1					
30	Zn	2	2	6	2	6	10	2					
31	Ga	2	2	6	2	6	10	2	1				
32	Ge	2	2	6	2	6	10	2	2				
33	As	2	2	6	2	6	10	2	3				
34	Se	2	2	6	2	6	10	2	4				
35	Br	2	2	6	2	6	10	2	5				
36	Kr	2	2	6	2	6	10	2	6				
37	Rb	2	2	6	2	6	10	2	6			1	
38	Sr	2	2	6	2	6	10	2	6			2	
39	Y	2	2	6	2	6	10	2	6	1		2	
40	Zr	2	2	6	2	6	10	2	6	2		2	

		N				O				P			Q
	…	4s	4p	4d	4f	5s	5p	5d	5f	6s	6p	6d	7s
55	Cs	2	6	10		2	6			1			
56	Ba	2	6	10		2	6			2			
57	La	2	6	10		2	6	1		2			
58	Ce	2	6	10	2	2	6			2			
59	Pr	2	6	10	3	2	6			2			
60	Nd	2	6	10	4	2	6			2			
61	Pm	2	6	10	5	2	6			2			
62	Sm	2	6	10	6	2	6			2			
63	Eu	2	6	10	7	2	6			2			
64	Gd	2	6	10	7	2	6	1		2			
65	Tb	2	6	10	9	2	6			2			
66	Dy	2	6	10	10	2	6			2			
67	Ho	2	6	10	11	2	6			2			
68	Er	2	6	10	12	2	6			2			
69	Tm	2	6	10	13	2	6			2			
70	Yb	2	6	10	14	2	6			2			
71	Lu	2	6	10	14	2	6	1		2			
72	Ho	2	6	10	14	2	6	2		2			
73	Ta	2	6	10	14	2	6	3		2			
74	W	2	6	10	14	2	6	4		2			
75	Re	2	6	10	14	2	6	5		2			
76	Os	2	6	10	14	2	6	6		2			
77	Ir	2	6	10	14	2	6	7		2			
78	Pt	2	6	10	14	2	6	9		1			
79	Au	2	6	10	14	2	6	10		1			
80	Hg	2	6	10	14	2	6	10		2			
81	Tl	2	6	10	14	2	6	10		2	1		
82	Pb	2	6	10	14	2	6	10		2	2		
83	Bi	2	6	10	14	2	6	10		2	3		
84	Po	2	6	10	14	2	6	10		2	4		
85	At	2	6	10	14	2	6	10		2	5		
86	Rn	2	6	10	14	2	6	10		2	6		
87	Fr	2	6	10	14	2	6	10		2	6		1
88	Ra	2	6	10	14	2	6	10		2	6		2
89	Ac	2	6	10	14	2	6	10		2	6	1	2
90	Th	2	6	10	14	2	6	10		2	6	2	2
91	Pa	2	6	10	14	2	6	10	2	2	6	1	2
92	U	2	6	10	14	2	6	10	3	2	6	1	2
93	Np	2	6	10	14	2	6	10	4	2	6	1	2
94	Pu	2	6	10	14	2	6	10	6	2	6		2
95	Am	2	6	10	14	2	6	10	7	2	6		2
96	Cm	2	6	10	14	2	6	10	7	2	6	1	2
97	Bk	2	6	10	14	2	6	10	9	2	6		2

		K	L		M			N				O	
		1s	2s	2p	3s	3p	3d	4s	4p	4d	4f	5s	5p
41	Nb	2	2	6	2	6	10	2	6	3		2	
42	Mo	2	2	6	2	6	10	2	6	4		2	
43	Te	2	2	6	2	6	10	2	6	5		2	
44	Rv	2	2	6	2	6	10	2	6	7		1	
45	Rh	2	2	6	2	6	10	2	6	8		1	
46	Pd	2	2	6	2	6	10	2	6	10			
47	Ag	2	2	6	2	6	10	2	6	10		1	
48	Cd	2	2	6	2	6	10	2	6	10		2	
49	In	2	2	6	2	6	10	2	6	10		2	1
50	Sn	2	2	6	2	6	10	2	6	10		2	2
51	Sb	2	2	6	2	6	10	2	6	10		2	3
52	Te	2	2	6	2	6	10	2	6	10		2	4
53	I	2	2	6	2	6	10	2	6	10		2	5
54	Xe	2	2	6	2	6	10	2	6	10		2	6

		N				O				P			Q
	...	4s	4p	4d	4f	5s	5p	5d	5f	6s	6p	6d	7s
98	Cf	2	6	10	14	2	6	10	10	2	6		2
99	Es	2	6	10	14	2	6	10	11	2	6		2
100	Fm	2	6	10	14	2	6	10	12	2	6		2
101	Md	2	6	10	14	2	6	10	13	2	6		2
102	No	2	6	10	14	2	6	10	14	2	6		2
103	Lr	2	6	10	14	2	6	10	14	2	6	1	2

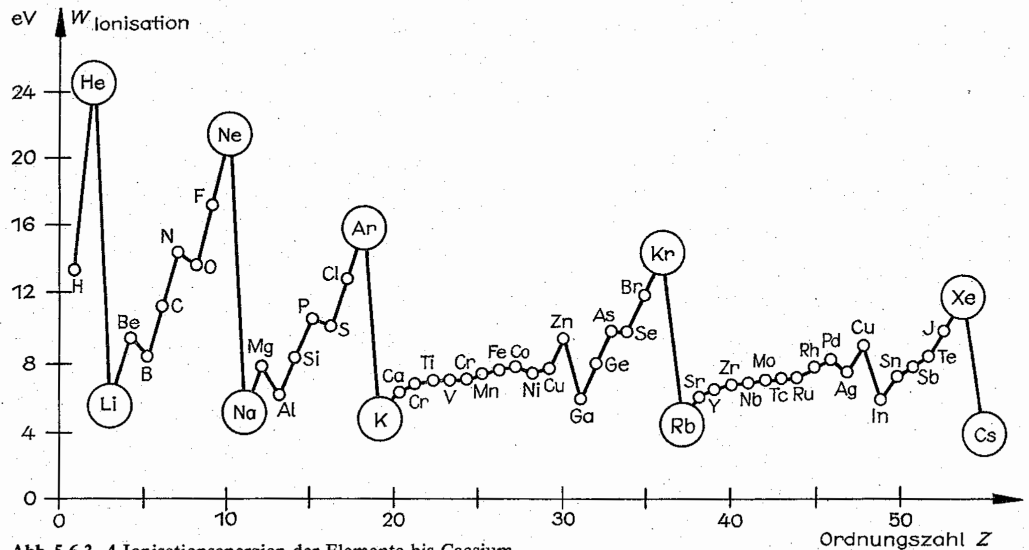


Abb. 5.6.3–4 Ionisationsenergien der Elemente bis Caesium

5.6.4. Das wellenmechanische Atommodell

5.6.4.1. Grenzen der Bohrschen Theorie. Den größten Erfolg erlangte das Bohrsche Atommodell mit der atomtheoretischen Erklärung des Periodensystems der Elemente. Fast zwanglos ergeben sich aus der Besetzung der „Planetenbahnen" mit Elektronen die Elemente. Jedoch zeigten sich an einigen Stellen schon die Grenzen der Theorie, aus denen vermutet werden konnte, daß der Atomismus eines eigenen theoretischen Gebäudes bedurfte. Da aber aus der Betrachtung der Atome auch alle makroskopischen Systeme und deren Eigenschaften folgen müssen, war für das gesamte Gebiet der Physik ein neues theoretisches Konzept zu erwarten.

Aus der experimentellen Spektroskopie konnte man entnehmen, daß für die Quantenzahlen, beispielsweise für den Bahndrehimpuls $[l(l+1)]^{1/2}$ statt l einzusetzen notwendig war. Auch die Existenz von Pendelbahnen mit dem Drehimpuls $l=0$ war nicht ableitbar, da das Elektron dabei den Kern durchdringen müßte. Überhaupt war die Quantelung der Drehimpulse zwar außerordentlich erfolgreich in der Auswahl der nur erlaubten Elektronenbahnen, aber in keiner Weise theoretisch fundiert. Besonders wenig sagte das Modell über die Strahlung bei Bahnsprüngen aus. Amplitude, Phase und Polarisationszustand sind nicht erklärbar. Ein auch nicht klärbares Problem war die Begründung der Größe des Eigendrehimpulses (Spin) des Elektrons. Der zugeordnete Drehimpuls verlangte für die Vorstellung einer rotierenden Kugel Umfangsgeschwindigkeiten größer als die Lichtgeschwindigkeit. Und selbst wenn der Drehimpuls akzeptiert werden könnte, ergibt er für kein Modell der Ladungsverteilung das richtige magnetische Moment.

Wesentliches Element und Handicap zugleich sind die Bohrschen Bahnen selbst. Spätestens mit Formulierung der Unbestimmtheitsbeziehung wird der Bahnbegriff problematisch. Angenommen, die Bahn sei auf einen Radiuswert von 10^{-12} m genau festgelegt, was schon zu fordern ist, will man von einer Bahn reden können, dann folgt aus der Unbestimmtheitsbeziehung eine Energieunschärfe, eine prinzipiell nicht genauer bestimmbare Schwankungsbreite der

Energie von $\approx 10^4$ eV. Damit wären die Energiewerte unakzeptabel undefiniert – um Größenordnungen über den eigentlichen Wert unscharf. Legt man andererseits die experimentellen Bindungsenergien auf 10 % fest, werden die Aufenthaltsorte der Elektronen über einen großen Radienbereich gleichermaßen möglich – der Begriff „Bahn" verliert seinen Sinn.

Das Planetenmodell ist anschaulich und wird über unsere Sinne vom Verstand gern angenommen. Die Quantentheorie, die in ungeahnter Perfektion alles Erklärbare auch erklärt und mit wenigen Grundannahmen auskommt, kann nicht mehr als anschaulich betrachtet werden. Typisch für diesen Zustand ist der Ausspruch eines Physikers, der formulierte: „Die Quantentheorie den Kleinstkindern". Nur einem von der normalen Umwelt noch nicht gebildeten Verstand ist ein Verständnis der Quantentheorie problemlos möglich. Ein Kompromiß ist empfehlenswert: Die Quantentheorie zu handhaben, den Formalismus zu gebrauchen und im Hinterkopf die Ergebnisse mit den Bohrschen Modellvorstellungen etwas zu unterlegen.

5.6.4.2. Die Grundaxiome der Quantentheorie. In diesem Abschnitt wird nur eine knappe Übersicht gegeben, um die Vorstellungen über das Atom aus quantentheoretischer Beschreibung erläutern zu können. Die Quantentheorie ist im Hauptkapitel 4 umfassend dargestellt.

Die Theorien der klassischen Physik werden um zwei Grundaxiome erweitert, es sind dies:

1. Der Welle-Teilchen-Dualismus. Er wird durch die de Broglie-Beziehung $p = h/\lambda$ beschrieben. Jedem bewegten klassischen Teilchen ist eine Welle der Wellenlänge λ zugeordnet, und jede Welle im klassischen Sinne ist mit einem Impuls p behaftet.
2. Es gilt die Energiequantelung gemäß der Einsteinschen Gleichung $W = h\nu$. Danach haben elementare Quanten den ihrer Frequenz ν proportionalen Energieinhalt W.

Die quantitative Verknüpfung der Meßgrößen p und λ bzw. W und ν erfolgt durch das Plancksche Wirkungsquantum h ($h = 6{,}6 \cdot 10^{-34}$ Js). Um die durch die Axiome gegebenen neuen Eigenschaften beschreiben zu können, wird sehr abstrakt eine Wellenfunktion Ψ (Psi-Funktion) eingeführt. Ihre Bedeutung ist noch zu diskutieren. Sie soll der allgemeinen Wellengleichung genügen: $\Delta\Psi = (1/u^2)\,\partial^2\Psi/\partial x^2$. Δ ist die Kurzschreibweise für den Laplace-Operator, $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$. Über u, die Phasengeschwindigkeit, können die beiden Grundaxiome in die Wellengleichung eingeführt werden, wenn man einsetzt $u = \lambda \cdot \nu = h\nu/mv$. Weiterhin ist die Gültigkeit des Erhaltungssatzes der Energie zu fordern. Die Gesamtenergie W eines Systems ist gleich der Summe aus potentieller Energie W_{pot} und kinetischer Energie W_{kin}, $W = mv^2/2 + W_{pot}$. Auch der Energiesatz kann über die Phasengeschwindigkeit durch $1/u^2 = [2\,m(W - W_{pot})]/(h^2\nu^2)$ in die Wellengleichung eingebracht werden. Daraus ergibt sich die Wellengleichung zu $\Delta\Psi = [2\,m(W - W_{pot})]/(h^2\nu^2) \cdot \partial^2\Psi/\partial t^2$. Für viele Probleme genügt es, nur den zeitunabhängigen Teil dieser Differentialgleichung zu betrachten. Diese Vereinfachung ist möglich, da die Ψ-Funktion sich als Produkt eines nur ortsabhängigen und eines nur zeitabhängigen Teils schreiben läßt: $\Psi(t, x, y, z) = \Psi(x, y, z)\exp(-2\pi i\nu t)$. Diesen Ansatz zweimal differenziert und eingesetzt, führt zu

$$\Delta\Psi + 2\,m/\hbar^2\,(W - W_{pot})\,\Psi = 0 \text{ (Schrödinger-Gleichung)}$$

bzw. $(\hbar^2/[2\,m]\,\Delta + W_{pot})\,\Psi = W\Psi$

oder noch kürzer $\quad \boldsymbol{H}\Psi = W\Psi$ ($\boldsymbol{H}$ = Hamilton-Operator).

Das ist die von SCHRÖDINGER eingeführte zeitfreie Differentialgleichung zur Beschreibung stationärer Zustände atomarer Systeme.

In der klassischen Physik besteht für die Lösung physikalischer Probleme die schwierigste Aufgabe in der Aufstellung der für den Einzelfall gültigen Gleichung. Ist erst die Gleichung aufgestellt, bereitet ihre Lösung kaum noch Probleme. In der Quantentheorie ist die Grundgleichung gegeben. Probleme bereitet ihre Lösung. Mitunter ist das wirkende Potential in seinem räumlichen Verlauf nur näherungsweise bekannt. Die Lösung der Wellengleichung erfordert es noch, Randwerte vorzugeben. Der Rechenaufwand bei der Suche nach Werten der Ψ-Funktion, die Lösungen der Differentialgleichung darstellen, kann erheblich sein.

Ein sehr einfaches, auch durch die Wellengleichung lösbares Problem ist das der schwingenden Saite. Als Randwerte bezeichnet man die durch Einspannen der Enden vorgegebenen Werte der Ψ-Funktion (bei einer Saite der Länge l wird bestimmt $\Psi[0] = 0$, $\Psi[l] = 0$). Als mögliche Lösungen ergeben sich stehende Wellen mit Schwingungsmaxima und -knoten. Ganz analog sind die durch die Drehimpulsquantelung $r_n p = nh$ gegebenen Bohrschen Bahnen (vgl. 5.6.1.3.) betrachtbar. Unter Verwendung der de Broglie-Beziehung sind die zugehörigen Wellenlängen λ gerade ganzzahlige Vielfache der Bahnlängen ($n \cdot \lambda = 2\pi r_n$). Das Wellenbild liefert demnach als mögliche Lösungen für stabil existierende Elektronenbahnen stehende Wellen.

5.6.4.3. Das Wasserstoffatom. Allgemein läßt sich zeigen, daß das gebundene Elektron nur gequantelte Energiewerte annehmen kann. Diese Energiewerte werden Energieeigenwerte genannt, die zugehörigen Funktionen Eigenfunktionen.

Zur Berechnung der Energiezustände des Wasserstoffatoms wird das Coulomb-Potential $W_{pot} = -e^2/(4\pi\varepsilon_0 r)$ eingesetzt. Die Rechnung erfolgt zweckmäßigerweise in Kugelkoordinaten [$\Psi(x, y, z) \rightarrow \Psi(r, \vartheta, \varphi)$]. Eine Aufspaltung der Differentialgleichung in drei jeweils nur von einer Koordinate abhängigen Differentialgleichungen wird möglich. Aus der Randbedingung des Verschwindens der Lösungsfunktion im Unendlichen (Normierbarkeit) entstehen aus dem Radialanteil die schon aus dem Bohrschen Modell bekannten Energiewerte

$W_n = -(1/n^2)\,(m_e e^4)/[8 \cdot (2\pi\varepsilon_0 h)^2]$, $n = 1, 2, 3, \ldots$

Die anderen Differentialgleichungen liefern eindeutige und stetige Lösungen nur für die Quantenzahlen $l = 0, 1, 2, 3, \ldots, n-1$, und $m = 0, \pm 1, \pm 2, \ldots, \pm l$.

Der Formalismus liefert nicht den Elektronenspin, nicht die Spinquantenzahl. Der Eigendrehimpuls des Elektrons muß als Elementarteilcheneigenschaft eingeführt werden.

Die vollständigen Eigenfunktionen des H-Atoms sind Kugelflächenfunktionen. Die Ψ-Funktion selbst ist keine Meßgröße, aber das Quadrat der Ψ-Funktion gibt eine Wahrscheinlichkeitsdichte an. Die Wahrscheinlichkeit, ein Elektron im Volumenteil $\mathrm{d}V$ zu finden, ist $\Psi^*\Psi\,\mathrm{d}V$. Infolgedessen sind die Eigenfunktionen zu normieren: $\int_{(V)} \Psi^*\Psi\,\mathrm{d}V = 1$, was bedeutet, daß das Elektron mit der Wahrscheinlichkeit 1 (also mit Sicherheit) irgendwo im betrachteten Volumen V zu finden ist. Das Quadrat der Kugelflächenfunktionen für das Wasserstoffatom gibt die *Aufenthaltswahrscheinlichkeiten* für das Elektron an. Je nach den Quantenzahlen des Zustandes, in dem es sich befindet, ergeben sich „verschmierte" Aufenthaltsgebiete. Im Bohrschen Atommodell sind die Bahnen genau festgelegt, zwi-

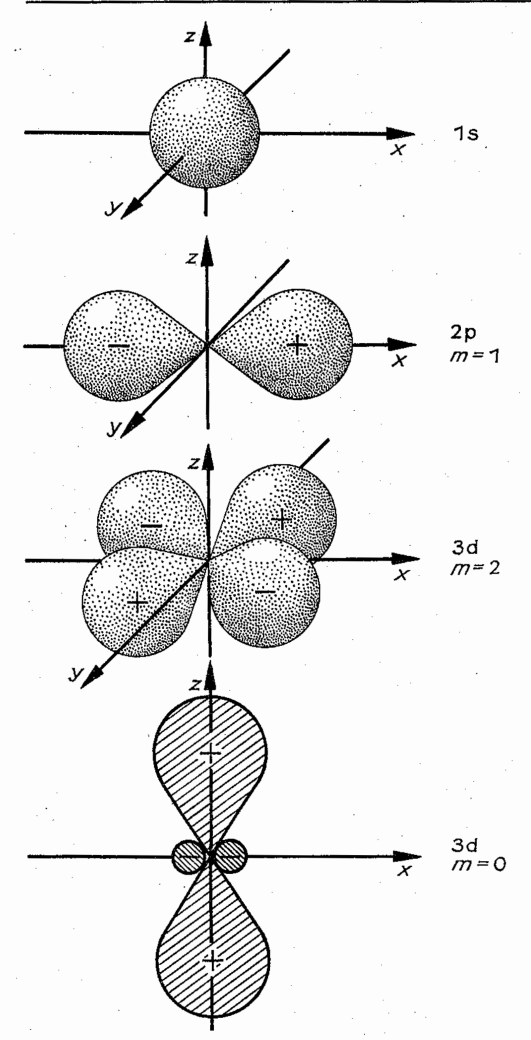


Abb. 5.6.4–1 Winkelverteilung der Elektronen für verschiedene Zustände

schen ihnen befinden sich ausgedehnte verbotene Gebiete für den Aufenthalt von Elektronen. Die Quantentheorie liefert Wahrscheinlichkeitsräume, große Gebiete mit einer Wahrscheinlichkeitsverteilung größer als Null. Der Begriff „Bahn" ist nicht mehr sinnvoll zu gebrauchen. Für den Bereich, in dem sich das Elektron überwiegend aufhält (mit 90 % Wahrscheinlichkeit), wurde der Begriff *Orbital* eingeführt. Die noch einfach darzustellenden Winkelabhängigkeiten der Orbitale der 1s-, 2p- und 3d-Zustände sind in Abb. 5.6.4–1 gezeigt. Der 1s-Zustand hat Kugelsymmetrie. Die 2p-Orbitale haben die Form einer Doppelkeule. Die drei Einstellmöglichkeiten der p-Zustände ergeben sich durch verschiedene Raumorientierung. Die Keulen sind in einer Einstellung rotationssymmetrisch um die x-Achse angeordnet, wie im Bild dargestellt. Die beiden anderen Möglichkeiten sind die rotationssymmetrisch zur y- bzw. zur z-Achse. Sind alle drei p-Orbitale mit Elektronen besetzt, ergibt sich eine recht gute Raumfüllung. Im Koordinatenursprungspunkt geht der winkelabhängige Anteil der Ψ-Funktion durch Null und ändert sein Vorzeichen. Das Vorzeichen ist für die Darstellung der Aufenthaltswahrscheinlichkeit unerheblich. Es ist aber mit eingezeichnet, weil es bei Molekülorbitalen eine Rolle spielt. Dort überlappen sich Anteile der Ψ-Funktionen verschiedener Atome. Je nach dem, ob sie gleichen oder verschiedenen Vorzeichens sind, entstehen bindende oder antibindende Orbitale. Auch die 3d-Orbitale haben noch Keulenform. Vier der fünf Einstellmöglichkeiten sind von der im Bild dargestellten Art, sie unterscheiden sich nur durch ihre Lage im Raum. Der fünfte Orbital ist rotationssymmetrisch zur z-Achse und mit der Bezeichnung $3d_{z^2}$ als Schnittbild gezeigt.

Es existiert noch eine andere Möglichkeit, die Form und Lage der Orbitale anschaulich zu machen. In Analogie zu den Knotenpunkten der stehenden Wellen einer schwingenden Saite bezeichnet man die Flächen verschwindender Aufenthaltswahrscheinlichkeit für das Elektron als Knotenflächen. Auf ihnen ist die Ψ-Funktion Null. Die Knotenflächen der s-Zustände sind Kugeln, die des 2p-Zustandes für $m = 0$ ist eine Ebene, und die des 3d-Zustandes für $m = 0$ besteht aus zwei auf die Spitze gestellten Kegelflächen.

Interessant im Vergleich zu den Bohrschen Bahnen sind die in Abb. 5.6.4–2 abgebildeten *Radialverteilungen* für die Aufenthaltswahrscheinlichkeiten im Wasserstoffatom. Über

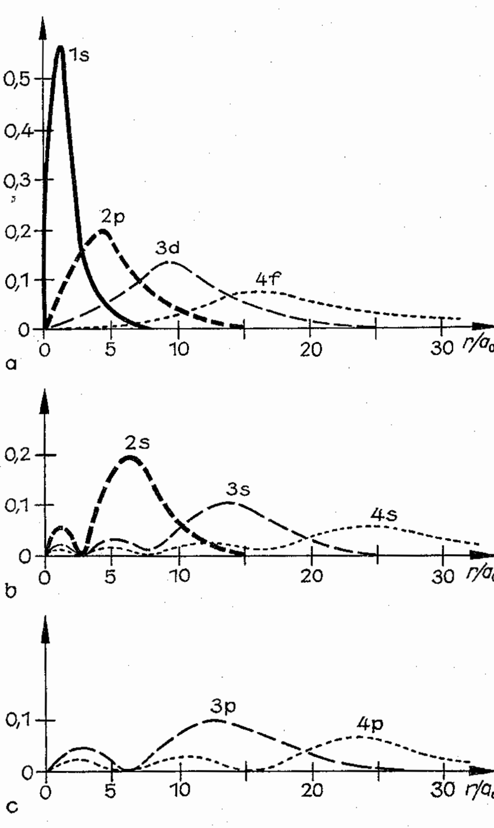


Abb. 5.6.4–2 Radiale Aufenthaltswahrscheinlichkeiten des Elektrons im Wasserstoffatom

den Radius R ist die Wahrscheinlichkeit, das Elektron zwischen r und $r + dr$ anzutreffen, aufgetragen. Die Kurven im Bildteil a gelten für die schon in ihrer Winkelabhängigkeit gezeichneten 1s-, 2p- und 3d-Orbitale. Die Wahrscheinlichkeitsverläufe der Bildteile b und c betreffen die s- bzw. p-Elektronen größerer Hauptquantenzahlen. r ist in Einheiten des Bohrschen Radius a_0 gegeben. Mit zunehmender Hauptquantenzahl werden die Orbitale breiter, verschmieren stärker. Die Verteilung hat aber immer noch Gebiete bevorzugten Aufenthaltes der Elektronen. Die größte radiale Aufenthaltswahrscheinlichkeit der 1s-, 2p-, 3d- und 4f-Elektronen findet man in Übereinstimmung mit der Lage der Bahnen nach BOHR (vgl. 5.6.1.3.) bei $1a_0$, $4a_0$, $9a_0$ bzw. $16a_0$. Die Abb. weist noch aus, daß die Elektronen zwar überwiegend in einem Orbital lokalisiert sind, aber auch noch eine geringe Wahrscheinlichkeit besitzen, sich in einem Zustand mit kleinerer Hauptquantenzahl n aufzuhalten.
Die Quantentheorie beschreibt nicht nur die stationären Elektronenzustände im Atom. Sie liefert Regeln für die erlaubten und verbotenen Übergänge. Sie erklärt auch die Auswahlregeln für die Polarisation der Strahlung. Die Übergangswahrscheinlichkeiten sind genau berechenbar.

5.6.5. Röntgen-Strahlen, Laser

5.6.5.1. Erzeugung von Röntgen-Strahlen. 1895 entdeckte KONRAD RÖNTGEN die nach ihm benannten Strahlen. In der englischsprachigen Literatur ist die von ihm vorgeschlagene Bezeichnung X-Strahlen gebräuchlich. *Röntgen-Strahlen* sind energiereiche Quanten elektromagnetischer Strahlung im Wellenlängenbereich zwischen ≈ 1 pm und 1 nm. Das entspricht dem Energiebereich von ≈ 1 MeV bis 1 keV. Zu niederen Energien hin schließt sich das Gebiet der harten UV-Strahlung an, bei Energien ab ≈ 100 keV spricht man auch von γ-Strahlen. Die relativ zum Licht große Energie der Röntgen-Strahlen bewirkt eine gute Durchdringungsfähigkeit. Zusammen mit den Eigenschaften, Stoffe zur Fluoreszenz anzuregen und fotografische Platten zu schwärzen, kam es recht schnell zu einer verbreiteten Anwendung des „Röntgens" in Medizin und Technik.
Den schematischen Aufbau einer Röntgen-Röhre zeigt Abb. 5.6.5–1. In einem evakuierten Kolben werden durch die Glühkatode Elektronen erzeugt und mittels hoher Spannungen zur Antikatode hin beschleunigt. Moderne Röntgen-Röhren arbeiten mit Beschleunigungsspannungen bis zu 300 kV. Beim Aufprall dort entsteht die Strahlung. Die Antikatode wird thermisch stark belastet, besteht deshalb aus hochschmelzbarem Metall (z.B. Wolfram) und wird bei höherer Leistung zusätzlich mit Wasser gekühlt. Auch Teilchenbeschleuniger der Kernphysik werden zur Erzeugung von Röntgen-Strahlen genutzt. Besonders dazu geeignet ist das Betatron (vgl. 5.4.3.2.). Die Energie dieser „Elektronenschleuder" liegt im 10-MeV-Bereich. Die damit erzeugbare sehr kurzwellige Röntgen-Strahlung ist besonders durchdringend.

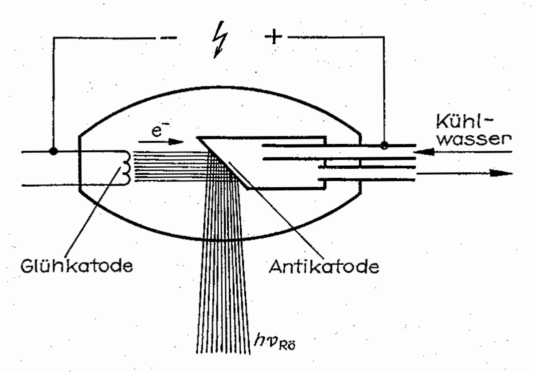


Abb. 5.6.5–1 Schematischer Aufbau einer Röhre zur Erzeugung von Röntgen-Strahlen

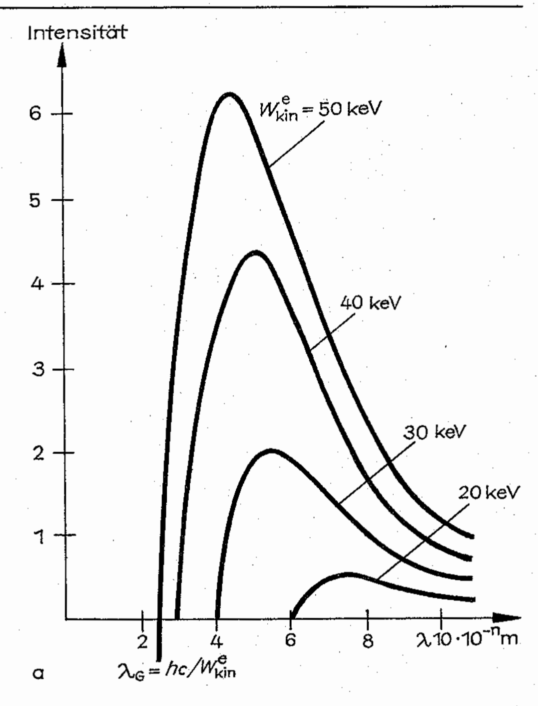
Röntgen-Strahlung besteht aus zwei Komponenten, einem Anteil mit kontinuierlichem Frequenzverlauf und intensiven Linien geringer Frequenzbreite. Die erste Komponente wird das *Röntgen-Bremskontinuum* genannt und enthält die überwiegende Strahlungsintensität. Die zweite Komponente besteht aus wenigen Linien, durch die das Katodenmaterial eindeutig charakterisiert ist, sie wird *charakteristische Röntgen-Strahlung* genannt. Beide Komponenten haben verschiedene Entstehungsursachen.
5.6.5.2. Charakteristische Röntgen-Strahlung. Die scharfen Linien im Röntgen-Spektrum entstehen durch Elektronenübergänge zwischen den inneren Niveaus der Atome. Die Bindungsenergien der K-Schalen-Elektronen sind sehr hoch im Vergleich zu denen äußerer Elektronen. Bindungsenergiedifferenzen von Elektronen auf den äußeren Bahnen, die zur Aussendung sichtbaren Lichtes führen, betragen wenige Elektronenvolt. Beim Übergang von Elektronen aus der L- in die K-Schale wird, überschlagsmäßig gerechnet, die Energie W (Kα) frei, $W(K_\alpha) = 13{,}6\,(Z-1)^2$ [eV]; Z = Ordnungszahl des betrachteten Elementes. Die Bezeichnung K_α kennzeichnet zum einen den Übergang in die angegebene Schale (K) und zum anderen mit α, β, γ den Übergang aus den in dieser Reihenfolge höheren Schalen. (K_α:L→K, K_β: M→K, K_γ: N→K) Die hohen Bindungsenergiewerte bis 100 keV erklären sich aus der sehr starken Anziehung der inneren Elektronen durch den Atomkern. Fast die gesamte Kernladung, die quadratisch in die Formel für die Bindungsenergie eingeht (vgl. 5.6.1.3.) wird wirksam. Betrachtet man ein Elektron der K-Schale, so wird die Kernladung nur durch das zweite Elektron der K-Schale um eine Einheit reduziert. Die in etwa lineare Abhängigkeit der Frequenz der K-Linien vom Quadrat der Ordnungszahl wurde schon 1913 von MOSELEY festgestellt. Sie war eine wesentliche Stütze zur Festlegung der Reihenfolge der Elemente im PSE. Heutzutage beruht eine wichtige Methode der Elementanalyse auf der Frequenzbestimmung der charakteristischen Röntgen-Strahlen. Die charakteristischen Röntgen-Strahlen sind intensitätsstark genug, um sie aus der Gesamtstrahlung herausfiltern und damit Strukturuntersuchungen vornehmen zu können. Auch der umgekehrte Prozeß ist von Bedeutung. Mit der charakteristischen Röntgen-Strahlung werden Elektronen auch kernnaher Bahnen aus den Atomen gelöst. Die Energiebestimmung der Elektronen gestattet Hinweise auf die Art der chemischen Bindung. Hierzu wird beispielsweise die Mg_{K_α}- oder die Al_{K_α}-charakteristische Linie verwendet, mit 1253,6 bzw. 1486,6 eV Energie. Am effektivsten werden Röntgen-Strahlen mit Elektronen angeregt. Die hochbeschleunigten Elektronen stoßen mit den Elektronen der Atome der Antikatode. Die Atome werden ionisiert, auch unter Entfernung von Elektronen innerer Schalen. Ist dadurch ein Platz beispielsweise in der K-Schale frei geworden, so geht sofort ein Elektron aus einer weiter außen liegenden Schale in diesen freien Zustand festerer Bindung über. Die Bindungsener-

giedifferenz wird in Form von Röntgen-Strahlen abgegeben. Entsprechend der Feinstrukturierung der Energieniveaus innerhalb einer Schale entsteht auch eine Aufspaltung der charakteristischen Röntgenlinien. Nicht jeder Übergang von einem Elektronenniveau zu allen tiefer liegenden ist möglich. Der Erhaltungssatz für den Drehimpuls führt zu der Auswahlregel, daß nur Übergänge mit einer Bahndrehimpulsänderung von $\Delta l = \pm 1$ bzw. Änderungen des Gesamtdrehimpulses $\Delta j = \pm 1, 0$ unter Emission eines Röntgen-Quantes möglich sind.

Auger-Effekt. Wenn die Bindungsenergie W_B beispielsweise eines Elektrons der K-Schale größer ist als die von zwei Elektronen einer anderen, beispielsweise der L-Schale ($W_B^K > 2\,W_B^L$), kann ein Auger-Elektron beobachtet werden. Entsteht durch Elektronenbeschuß eine Vakanz in der K-Schale, so wird ein Elektron der L-Schale den freien Platz einnehmen, und die Bindungsenergiedifferenz kann auf ein anderes Elektron der L-Schale übertragen werden. Dieses Elektron verläßt dann das Atom mit der kinetischen Energie $W_{kin} = W_B^K - 2\,W_B^L$, es wird nach dem Entdecker dieses Effektes *Auger-Elektron* genannt. Das Auger-Elektron übernimmt mitunter nicht die gesamte Energiedifferenz. In den Folgeprozessen des „Nachrückens" von Elektronen in die in diesem Beispiel freiwerdenden Zustände der L-Schale kann noch ein bzw. können weitere Auger-Elektronen emittiert werden. Da die kinetischen Energien der Auger-Elektronen nur von den Bindungsenergien des emittierenden Atomes abhängen, charakterisieren sie dieses Atom. Die Spektroskopie von Auger-Elektronen ist eine aktuelle Methode zur Materialanalyse. Da die Auger-Elektronen im Inneren eines Materials schnell ihre Energie verlieren, sind Auger-Spektrometer Geräte speziell zur Oberflächenanalyse.

Röntgen-Absorptionsspektren. Die bisher behandelten Röntgen-Linien sind Emissionslinien – Abstrahlungen der Atome. Bestrahlt man Atome mit Röntgen-Licht, so werden auch Absorptionslinien beobachtet. Im Unterschied zur optischen Spektroskopie, wo Emissions- und Absorptionsspektren relativ gleichartig erscheinen, unterscheiden sich beide Fälle im Bereich der Röntgen-Strahlen erheblich. Emissionslinien erscheinen, wie besprochen, durch das Auffüllen von Elektronenleerstellen innerer Bahnen. Die Absorption von Röntgen-Strahlen erfolgt an Atomen, die keine leeren inneren Niveaus haben (sich im Grundzustand befinden). Der Übergang von Elektronen innerer Schalen durch Absorption von Röntgen-Quanten kann deshalb nur in freie Zustände erfolgen. Dort liegen die Zustände sehr dicht und gehen in das Kontinuum der freien Elektronen über. Es ergeben sich dadurch keine Absorptionslinien, sondern nur Absorptionskanten. Auch erscheinen zwar alle Zustände durch Übergänge in freie äußere Bahnen, aber alle Übergänge, die in besetzte Zustände erfolgen müßten, erfolgen nicht, so daß in der Regel Emissionslinien und Absorptionskanten nicht übereinstimmen können.

5.6.5.3. Röntgen-Bremsstrahlung. Das Bremskontinuum liefert den Hauptanteil der Strahlung. Je höher die Elektronenenergie, um so effektiver ist die Anregung. Eine charakteristische Größe der Energieverteilung ist die kurzwellige Grenze λ_G. Sie entspricht einer oberen Grenze für die Energie der Röntgen-Quanten. Quanten dieser Grenzenergie hc/λ_G entstehen unter Übernahme der gesamten kinetischen Energie eines auf die Antikatode beschleunigten Elektrons ($W_{kin}^e = eU$; U = Beschleunigungsspannung). Bei der Entstehung von Quanten geringerer Energie (längerer Wellenlänge) werden die Elektronen nur teilweise abge-



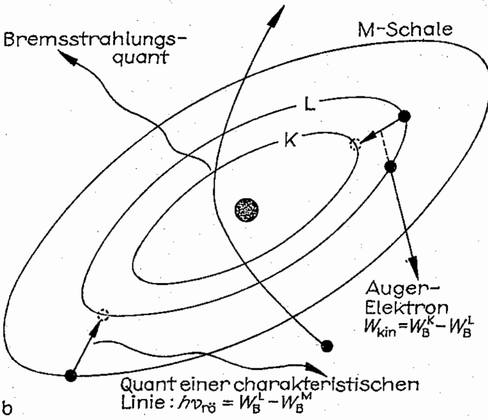


Abb. 5.6.5–2 (a) Röntgen-Bremsstrahlungskontinuum, (b) Entstehung der Röntgen-Strahlungskomponenten

bremst. Quanten aller Energiewerte bis zur Grenzenergie werden erzeugt. Die Quantenerzeugung gehorcht der Notwendigkeit, die im Abbremsvorgang frei werdende Energie abzutransportieren. Sie kann auch als notwendig für den Übergang zwischen Bohrschen Bahnen angesehen werden. Die Bahnen liegen dann weit außerhalb der gebundenen Zustände des bremsenden Atoms im „Linienkontinuum" der freien Elektronen. Die Atome liefern das bremsende elektrische Feld, ohne selbst einen wesentlichen Energiebetrag zu übernehmen. Die Form der Intensitätsverteilung in Abhängigkeit von der Wellenlänge verdeutlicht Abb. 5.6.5–2a. Das Bild enthält als Parameter für die einzelnen Intensitätsverteilungen die Energie der die Strah-

lung erzeugenden Elektronen. Abb. 5.6.5–2b zeigt zusammenfassend die Effekte zur Erzeugung von Röntgen-Strahlen.

Röntgen-Computertomografie. Das Verständnis der Methode des Röntgens und das Wissen um die Anwendung in Medizin und Technik gehören weitgehend zum Allgemeinwissen. Ein neuer, sehr aussichtsreicher Einsatz mit Modellcharakter für andere Analysenmethoden wird durch die Computertomografie realisiert. Tomografie ist vom Sinn des Wortes her ein Röntgen-Schnittverfahren, beinhaltet die Darstellung einzelner Objekt*ebenen* auf dem Röntgen-Bild. Die *Computertomografie* mit Röntgen-Strahlen (sowie allgemein auch für andere Analysemethoden) ermöglicht durch den Einsatz hocheffektiver Rechentechnik eine – bis vor wenigen Jahren nur theoretisch erwogene – räumliche Darstellung durchstrahlter Objekte.

Wenn ein Arzt früher, hinter dem Leuchtschirm stehend, bei der Betrachtung des Thoraxes den Patienten bat, sich zu drehen, so wollte er einen räumlichen Eindruck trotz des ebenen Schirmbildes erhalten. Das Gedächtnis speichert die Bilder bei verschiedener Durchstrahlungsrichtung und erzeugt sich eine räumliche Vorstellung, etwa von der Lage der Rippen.

Analog arbeitet die Computertomografie. Ein Objekt wird aus verschiedenen Richtungen durchstrahlt. Die Bildwerte aus jeder Richtung, von modernen Detektorbauelementen aufgenommen, gelangen in den großen Speicher eines schnellen Rechners, der dann das räumliche Bild konstruiert und in beliebiger Form ausgeben kann. Derzeit sind die technischen Möglichkeiten noch nicht so, daß die Signale eines Röntgen-Strahles mit großer Querschnittsfläche aufgenommen werden können. Folgendes Verfahren, eine Schnittfläche in bestimmter Ebene durch den zu untersuchenden Körper darzustellen, wird angewendet (Abb. 5.6.5–3). Ein nur in einer Ebene sich auffächernder Röntgen-Strahl schwenkt um den zu untersuchenden Körper. Mit einer Zeile dicht an dicht liegender Röntgen-Strahldetektoren werden auf der gegenüberliegenden Seite die Strahlintensitäten registriert. Abtastschritte mit einem Schwenk der Röntgen-Röhre um nur wenige Grad sind üblich, so daß bei voller Umdrehung ≈ 100 Meßwerte der beispielsweise 100 Detektoren aufgenommen werden. Die Detektoren sind sehr empfindlich. Sie ermöglichen grafische Darstellungen, die noch wesentlich genauer abbilden, als es mit den ≈ 30 Graustufen eines Fernsehbildschirmes möglich ist.

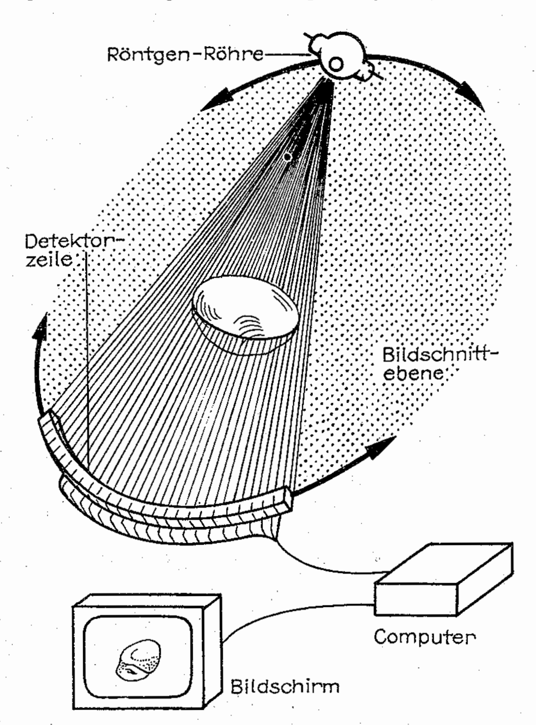


Abb. 5.6.5–3 Funktionsprinzip eines Röntgen-Computertomografen

Je größer die Ortsauflösung ist und je genauer die Meßwerte für das durchstrahlte Gewebe aufgenommen werden, um so eher erreicht man die wirklichkeitsgetreue Abbildung einer Körperschicht. Die Ortsauflösung liegt für Anwendungen in der Medizin bei ungefähr 1 mm². Die Empfindlichkeit der Detektoren entscheidet auch über die Meßzeit. Heutzutage ist die Strahlenbelastung des lebenden Organismus ausreichend gering, problematisch ist noch die Abbildung sich bewegender Objekte.

5.6.5.4. Das Laserprinzip. *Laser* lassen sich gut mit den bisher dargestellten Modellvorstellungen zum Atombau erklären. Und mit dem Verständnis der Funktionsweise der Laser vertieft sich das Verständnis über den Zusammenhang zwischen Atombau und Spektrallinien. Laser ist ein Kunstwort aus den Anfangsbuchstaben von engl. *l*ightwave *a*mplification by *s*timulated *e*mission of *r*adiation (Lichtverstärkung durch induzierte Emission von Strahlung). Im Bereich der Mikrowellen läßt sich der gleiche Effekt realisieren, man gebraucht dafür den Ausdruck Maser (microwave ...).

Induzierte Emission. Normalerweise wird bei der Anregung eines Atoms ein Elektron auf ein weiter außen liegendes Niveau gehoben und kehrt spontan in den Grundzustand zurück. Die Rückkehr erfolgt innerhalb von $\approx 10^{-8}$ s unter Aussendung eines Lichtquantes. Die Anregung kann durch die Absorption eines Lichtquantes passender Energie erfolgen – wie sie auch erfolgt, sie ist immer erzwungen, induziert. Auch die Emission kann erzwungen werden. Bei der Existenz zweier Niveaus und der Einstrahlung von Lichtquanten mit einer Energie, die gleich der Differenz zwischen den zwei Niveaus ist, sind induzierte Emissision wie induzierte Absorption gleich wahrscheinlich. Dadurch ist bei Stimulierung die Zahl der Übergänge in der Zeiteinheit proportional der Zahl der Elektronen in den Ausgangszuständen. Die Absorption bei Lichteinstrahlung überwiegt, wenn der untere Zustand stärker mit Elektronen besetzt ist und die Emission, wenn der obere Zustand mehr Elektronen trägt. Ein Laser kommt dadurch in Betrieb, daß die induzierte Emission die Absorption zwischen zwei Niveaus weit überwiegt. Voraussetzung dafür sind ein stark mit Elektronen aufgefüllter oberer Zustand und ein gering besetzter unterer Zustand (Grundzustand). Diese Umkehr der normalen Besetzungsverhältnisse wird als *Besetzungsinversion* bezeichnet.

Zur Realisierung einer Besetzungsinversion sind mindestens drei Niveaus notwendig, das ist aus Abb. 5.6.5–4 ersichtlich. Das Grundzustandsniveau liefert die Elektronen, die zum Lasern auf dem Laserniveau benötigt werden. Um dort die Besetzungsinversion zu erwirken, werden die Elektronen erst mittels einer Lichtquelle in ein anderes Niveau

oder Band „gepumpt“, was deshalb als Pumpniveau bezeichnet wird (optisches Pumpen). Aus dem Pumpniveau gelangen die Elektronen über sehr schnell ablaufende Prozesse auf das Laserniveau. Dadurch stellt sich zwischen Grundzustand und Pumpniveau kein Gleichgewicht in der Besetzung ein. Das Pumplicht, beispielsweise einer Xenon-Blitzlampe, bewirkt eine starke Entvölkerung des Grundzustandes. Aus dem Pumpniveau wechseln die Elektronen so schnell auf das Laserniveau, daß die stimulierte Rückkehr

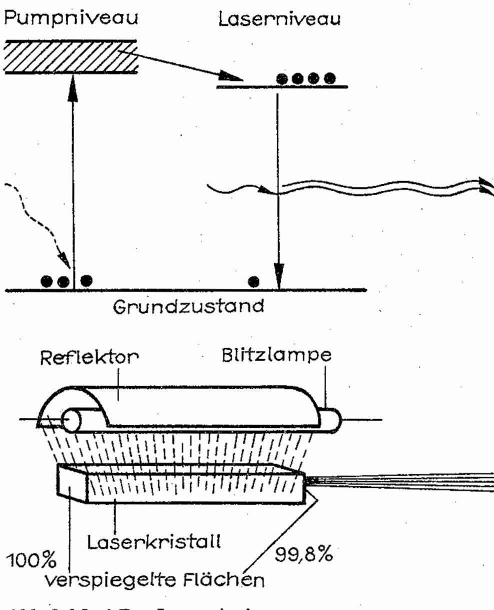


Abb. 5.6.5–4 Das Laserprinzip

nicht zustande kommt. Bei Gaslasern ist die Besetzungsinversion auch anders erreichbar. Mittels einer Gasentladung werden Ionen erzeugt, erhalten kinetische Energie, und durch Stöße untereinander werden die Gasteilchen auf das Pumpniveau angeregt. Die Gasart ist so gewählt, daß Übergänge aus dem Pumpniveau zurück in den Grundzustand optisch verboten sind. Elektronen auf derartigen Niveaus befinden sich in metastabilen Zuständen. Ihre Lebensdauer dort ist wesentlich länger, als wenn tieferliegende Niveaus unter Aussendung eines Lichtquantes erreichbar sind. Elektronen in metastabilen Zuständen haben Lebensdauern um 10^{-3} s. Der Elektronenübergang in das Laserniveau erfolgt strahlungslos. Diese strahlungslosen Übergänge in das nur etwas tiefer liegende Niveau sind Relaxationsprozesse. Beispielsweise durch Stoßprozesse in Gasen kann innerhalb von 10^{-10} s die Anregung aus den Pumpniveaus aller Atome auf die Laserniveaus übergehen.

Mit den aufgefüllten Laserniveaus existiert ein emissionsbereiter Zustand. Wird ein Lichtquant des Laserüberganges emittiert, kann es in einem anderen Atom den gleichen Übergang stimulieren. Diese nun zwei Lichtquanten können zwei weitere erzeugen, und der Prozeß wird lawinenartig anwachsen. Da die stimulierte Emission aber nicht sehr wahrscheinlich ist, zwingt man die Laserlichtquanten über Vielfachreflexionen in Laserresonatoren, sehr lange Wege in der Lasersubstanz zurückzulegen.

Laserresonatoren. Als Material für Festkörperlaser eignet sich ein chromiumdotierter Rubinkristall. Werden die Endflächen eines Rubinstabes so stark verspiegelt, daß fast alles Licht im Innern wieder reflektiert wird, wird das Licht oft hin und her laufen, ehe es den Kristall verläßt. Notwendig ist, daß die Spiegelflächen genau parallel zueinander stehen. Die Resonatoren von Gaslasern bestehen aus zwei beispielsweise im Abstand von 1 m angeordneten Spiegeln, zwischen denen in einem Glasrohr das laseraktive Gas (z.B. ein He-Ne-Gasgemisch) angeordnet ist. Die Verspiegelung eines Spiegels sollte möglichst 100 % sein. Durch den anderen Spiegel, mit $\approx 0{,}1$ % Transparenz, tritt das Laserlicht aus.

Es gibt Laser, die kontinuierlich Licht abstrahlen, und solche, die im Impulsbetrieb arbeiten. In „gütegeschalteten“ Lasern kann der Impuls extrem verkürzt werden, wodurch besonders hohe Leistungsdichten erreichbar sind. Für die Güteschaltung ist ein Spiegel zuerst verkippt und wird, den lawinenartigen Prozeß verzögernd, erst bei extremer Besetzungsinversion ausgerichtet. Mit Farbstofflasern läßt sich die Laserfrequenz über ein größeres Gebiet variieren. Um höhere Photonenenergien zu erreichen, sind nun auch im UV-Bereich arbeitende sogenannte Excimerlaser entwickelt worden. Kurzwellige Laser sind technisch schwer realisierbar, da die mit der stimulierten Emission konkurrierende spontane Emission mit der dritten Potenz der Frequenz anwächst.

5.6.5.5. Eigenschaften der Laserstrahlung. Monochromasie (Einfarbigkeit) bezeichnet die Eigenschaft einer Strahlung, von geringer Frequenzbreite zu sein. Monochromatisches Licht ist im vereinfachenden Sprachgebrauch Licht *einer* Wellenlänge, wenn auch die Unbestimmtheitsbeziehung eine natürliche Linienbreite fordert (vgl. 5.5.4.4.). Die Laserlichtquanten entstammen einem Übergang. Da alle Atome einer Art sehr gleich sind, sind auch alle Lichtquanten aus diesem Übergang der einzelnen Atome genau gleich. Die Niveaus können sehr schmal sein, relative Linienbreiten von $\Delta\nu/\nu \sim 10^{-14}$ sind beobachtbar. Laserfrequenzen können so genau stabil gehalten werden, daß mit ihnen die derzeitige Längeneinheit, das Meter, um zwei Größenordnungen genauer, auf zehn Stellen genau, festgelegt werden könnte. Mit Laser sind Entfernungen zum Mond auf ≈ 10 cm genau ausgemessen worden. Diesen Messungen kommt die durch die Resonatoranordnung bedingte hohe Parallelität des Lichtes zugute. Öffnungswinkel von nur 10^{-8} sr sind möglich. Aus dem Orbit werden Objekte auf der Erdoberfläche mit einer Genauigkeit von ± 1 m geortet. Die Monochromasie gestattet eine extrem scharfe Fokussierung des Lichtes. Dadurch kann man bohren, trennen, schweißen. Im Extremfall ist die zur Kernfusion ausreichende Energiedichte erzeugbar (Laserfusionsexperimente, vgl. 5.3.6.3.).

Kohärenz ist die hervorragendste Eigenschaft des Laserlichtes. Quanten der spontanen Emission haben eine zufallsbedingte Phasenlage, jedes Quant schwingt für sich. Die stimulierte Emission erzeugt Licht gleicher Phase. Die Quanten ordnen sich so, daß sie einen unendlich langen Wellenzug bilden. Derartiges Licht einer Phasenlage ist kohärent. Kohärentes Licht ist interferenzfähig. Mit Entwicklung der Laser wurde die optische Holografie möglich (vgl. 4.11.1.5.). Mittels des Laserlichtes sind auch wesentliche Fortschritte in der Nachrichtenübertragung zu erwarten. Den im Vergleich zur derzeit üblichen Nachrichtentechnik sehr hohen Frequenzen des Lichtes lassen sich extrem große Datenmengen aufmodulieren.

5.7. Aufbau von Molekülen

5.7.1. Allgemeine Eigenschaften

5.7.1.1. Der Molekülbegriff. Eine feste, zeitstabile Verbindung von zwei oder mehr Atomen wird Molekül genannt. Moleküle sind wie Atome im Grundzustand elektrisch neutral. Die Stabilität ist so zu verstehen, daß Energie aufgebracht werden muß, um eine Zerlegung in die Atome zu erreichen. Es muß die „Atomisierungsenergie" von außen zugeführt werden. Diese Energie wird bei der Bildung des Moleküls an die Umgebung abgegeben. Bekanntes Beispiel dafür ist die Verbrennung von Kohlenstoff unter Anwesenheit von Sauerstoff zu Kohlendioxid mit Freisetzung einer Bildungsenthalpie von $\Delta H = 394$ kJ mol^{-1}.
Charakteristisch für Moleküle ist die Kombination der Atome in nur wenigen möglichen und immer konstanten Proportionen. Bei den zweiatomigen Molekülen beispielsweise existiert H_2 oder NaCl. In der Natur findet man kein H_3 oder $NaCl_2$ oder andere Kombinationen. Atomarer Sauerstoff (O) existiert nicht als Gas, aber das zwei- und dreiatomige Molekül O_2 und O_3 (Ozon). Selbst wenn viele Proportionen möglich sind, sind sie konstant, wie zwischen Stickstoff und Sauerstoff: N_2O, NO, N_2O_3, NO_2, N_2O_5. Es können sich auch Moleküle aneinander legen, binden, aber die Bindungskräfte sind wesentlich schwächer als die innerhalb des Moleküls und erst für den Übergang Molekül-Festkörper von Bedeutung. Die Stabilität der Moleküle kann sehr verschieden sein. Bei tiefen Temperaturen genügen extrem geringe Bindungsenergien zur stabilen Existenz eines Moleküls. Im Bereich der Zimmertemperatur können einige schon thermisch instabil sein und an ihren eigenen Schwingungen zerbrechen. Andere zerfallen bei geringer Energiezufuhr durch Stoß oder Oxidation an Luft. Prinzipiell sind alle Moleküle in ihre Atome zerlegbar. Viele Moleküle, wie H_2O oder CO_2, sind unter Normalbedingungen stabil.
Die Molekülbindungen beruhen auf der Wirkung elektrischer Ladungen. Daß sich, trotz der von der Physik her gut beherrschten Kräfte, Begriffe wie „chemische Bindung", „chemische Bindungskräfte" ausbilden konnten, liegt an der Besonderheit des Vorliegens gerichteter Bindungen. Die Kräfte wirken nur in bestimmte Richtungen, wodurch sich beispielsweise konstante Bindungswinkel ausbilden. Die beiden Wasserstoffatome im H_2O bilden zum Sauerstoff einen Winkel von 105°.
Die Physik der Moleküle existiert harmonisch neben der Molekülchemie. Chemiker und Physiker untersuchen mit den ihnen eigenen Methoden das gleiche Objekt. Der Grad der Verflechtung ist groß, wofür auch die Existenz bedeutender Institute mit den Bezeichnungen Physikalische Chemie und Chemische Physik Zeugnis ablegen. Gemeinsam ist den beiden Grundlagenwissenschaften die Theorie über den Bau der Moleküle – erste und ideale Verständigungsbasis. Diese Theorie ist aus der über den Atombau entstanden. So, wie die Quantentheorie ein vertieftes Verständnis und die Beschreibung des Atombaues ermöglichte, ermöglichte sie in gleicher Weise Verständnis und Beschreibung der Moleküle.

5.7.1.2. Einige physikalische Eigenschaften. Molekülgröße. Sie ergibt sich oft aus der der Atome. Die Atome ordnen sich, dicht an dicht liegend, zu Molekülen. Dadurch ist die Molekülgröße durch die Größe und Anzahl ihrer Atome gegeben. Zweiatomige Moleküle haben einen Durchmesser von ≈ 3 bis $4 \cdot 10^{-10}$ m. Fadenförmige Makromoleküle haben Längen von 10^{-6} m, und die gestreckte Länge des DNS-Moleküls erreicht 1 mm.
Die Größe einfacher Moleküle läßt sich schon aus der van der Waalsschen Zustandsgleichung realer Gase abschätzen: In $(p + a/V^2) \cdot (V - b) = n\,RT$ ist das Volumen des Gases V um das Kovolumen b reduziert. Dieses entspricht dem vierfachen Molekülvolumen (± 30 %). Die innere Reibung bei Gasen hängt von der mittleren freien Weglänge λ der Moleküle ab. Aus der mittleren freien Weglänge läßt sich der Molekülradius r bestimmen, $1/\lambda = 4\sqrt{2}\,N r^2$ (N = Gasdichte, r = Molekülradius).
Moderne Methoden, Atomabstände zu bestimmen, nutzen die Röntgen-Strahl- oder Elektronenbeugung (vgl. Elektronenoptik, Elektronenmikroskop).
Dipolmomente sind von besonders anschaulicher Aussagekraft für zweiatomige Moleküle. Man unterscheidet zwischen permanenten, ohne äußere Einwirkung vorhandenen Dipolmomenten und induzierten, durch äußere Einwirkungen hervorgerufenen Dipolmomenten. Permanente Dipolmomente entstehen durch Auseinanderrücken der positiven und negativen Ladungsschwerpunkte. Das Dipolmoment $\boldsymbol{M}$ ist das Produkt aus dem Abstand l dieser Ladungsschwerpunkte und der Größe der verschobenen Ladung q, $M_p = l \cdot q$ (*permanentes Dipolmoment*).
Die Dipolmomente gleichatomiger Moleküle, z. B. von O_2, N_2 oder Cl_2 sind Null. Die Gleichgewichtsabstände r_0 der Atome betragen in den drei Fällen $r_0 = 1{,}20$, 1,10 bzw. $1{,}98 \cdot 10^{-10}$ m. Offensichtlich bilden sich die bindungsbedingten Veränderungen so aus, daß die Gesamtschwerpunkte der positiven und negativen Ladungen zusammenfallen. Andere Moleküle haben ein großes permanentes Dipolmoment, bis in die Größenordnung $r_0 \cdot e_0$ (e_0 = Elektronenladung). Bei derartigen Molekülen kann man vereinfachend annehmen, daß ein Elektron das Atom gewechselt hat, wodurch aus den beiden Molekülatomen ein positives und ein negatives Ion entstand. Typische Vertreter derartiger Ionenmoleküle sind beispielsweise NaCl, KCl und HF. Kaliumchlorid hat ein Dipolmoment von $e_0 \cdot 1{,}3 \cdot 10^{-10}$ m. Das entspricht der Vorstellung eines Moleküls aus zwei Ionen, da der Abstand der Ladungsschwerpunkte mit dem Abstand der Ionen vergleichbar ist. Beide angeführten Beispiele sind Grenzfälle. Zum einen realisiert sich selten das reine Ionenmolekül durch vollständigen Elektronenwechsel, zum anderen wirkt mitunter dem Auseinanderrücken der Ladungsschwerpunkte eine Ladungsverschiebung innerhalb der Atome entgegen. Diese Polarisation der Atome verringert dann das Dipolmoment.
Bei Molekülen aus mehr als zwei Atomen kann auf die Anordnung der Atome geschlossen werden. CO_2 hat kein permanentes Dipolmoment. Dadurch ist nahegelegt, daß es eine gestreckte Form mit dem C-Atom in der Mitte ausbildet. Für das H_2O wird ein Dipolmoment von $e_0 \cdot 0{,}38 \cdot 10^{-10}$ m gemessen, wodurch eine geknickte Form angenommen werden kann. Das sehr symmetrische Tetrachlorkohlenstoffmolekül (CCl_4) hat kein permanentes Dipolmoment. Vier Chloratome bilden die Ecken eines Tetraeders, in dessen Zentrum das Kohlenstoffatom angeordnet ist.
Induzierte Dipolmomente entstehen durch Auseinanderrücken der Ladungsschwerpunkte unter der Wirkung eines äußeren elektrischen Feldes. Das äußere elektrische Feld kann durch ein benachbartes Ion erzeugt sein. Die Verschiebung der negativen und positiven Ladungen ist dem äußeren Feld entgegengerichtet, das induzierte Dipolmoment $\boldsymbol{M}_i$ ist proportional der elektrischen Feldstärke $\boldsymbol{E}$, der

Proportionalitätsfaktor α wird als *Polarisierbarkeit* bezeichnet, $M_i = \alpha E$.

Die Polarisierbarkeit ist ein Maß für die Deformierbarkeit der Moleküle, hat die Dimension Kubikmeter und erreicht Werte in der Größenordnung der Volumina der Moleküle. Da diese selten kugelsymmetrisch sind, wird α in der Regel ein Tensor sein. Die richtungsabhängige Polarisierbarkeit eignet sich auch gut zur Aufklärung der Atomanordnungen im Molekül. Um den Polarisierbarkeitstensor ausmessen zu können, ist es notwendig, die Moleküle räumlich, etwa durch den orientierten Einbau in Kristalle, zu fixieren. Die Polarisierbarkeit des schon erwähnten CCl_4-Moleküls ist isotrop und der α-Wert in den drei Richtungen eines Koordinatensystems gleich groß. Auch das Helium, als Gas einatomig vorliegend (kein Molekül bildend), ist von isotroper Polarisierbarkeit. Anders im Falle des Cl_2. Hier ist die Polarisierbarkeit in Achsenrichtung fast doppelt so groß wie in Richtung der Koordinaten senkrecht dazu. Die Polarisierbarkeit wie auch das permanente Dipolmoment können mit verschiedenen Methoden ermittelt werden; etwa durch Lichtstreuung, durch den elektrooptischen Kerr-Effekt oder durch Messung der temperaturabhängigen Dielektrizitätskonstanten ε.

Messung von Dipolmomenten. Das mittlere permanente Dipolmoment $\bar{M}_p$, wie auch das induzierte Dipolmoment M_i, bilden als Produkt mit der Moleküldichte N die dielektrische Polarisation, $P = P_p + P_i$, $P_p = \bar{M}_p \cdot N$, $P_i = M_i \cdot N$. Sie bewirkt die dielektrische Verschiebung D des zu untersuchenden Molekülgases unter der Wirkung eines elektrischen Feldes der Feldstärke E, $D = E + 4\Pi P$, $D = \varepsilon E$, d.h. $\varepsilon = 1 + 4\Pi P/E$ (ε = Dielektrizitätskonstante). Das mittlere permanente Dipolmoment stellt sich gegen die Temperaturbewegung ein, $\bar{M}_p = M_p E M_p/(3kT)$, so daß sich eine temperaturabhängige Dielektrizitätskonstante ergibt: $\varepsilon = 1 + 4\Pi N (M_p^2/3kT + \alpha)$. Mißt man ε in Abhängigkeit von der Temperatur, so sind der Betrag des permanenten Dipolmomentes M_p und die Polarisierbarkeit α bestimmbar. Die Dielektrizitätskonstante ε ist aus dem Verhältnis der Kapazität eines Kondensators mit und ohne Dielektrikum zu ermitteln. ε ist definitionsgemäß gegeben zu $\varepsilon = C/C_{Vakuum}$ (vgl. 2.6.1.). Das Dielektrikum wäre in unserem Falle ein zu untersuchendes Molekülgas.

5.7.2. Chemische Bindung

Wie für den Bau der Atome, so gibt es auch für die Molekülbildung klassische Modelle. Sie existieren immer noch, obwohl eine geschlossene quantenmechanische Betrachtungsweise alles umfassender, manches erst richtig oder überhaupt erst erklärt. Für das Atom ist das Bohrsche Atommodell gut geeignet, Anschauungen zu vermitteln. Wegen der Anschaulichkeit nimmt man Unstimmigkeiten zur Realität in Kauf. Für das Molekül gilt Analoges. Die klassischen Betrachtungsweisen basieren auf dem Schalenmodell des Atoms. Acht Elektronen bilden eine abgeschlossene Schale. In ihr sind alle Zustände mit dem Bahndrehimpuls $l = 0$ und $l = 1$ besetzt (2 s- und 6 p-Elektronen). In der Außenschale diese Edelgaskonfiguration zu erreichen, ist das Optimierungsprinzip für den Bau der Moleküle. Mit Ausnahme des Heliums ist es eine Achterkonfiguration. Die Betrachtungen werden auch als *Oktett-Theorie* bezeichnet. Die im folgenden zu besprechenden Bindungsarten erreichen die Edelgaskonfiguration auf verschiedenen Wegen. Man unterscheidet zwischen 1. der kovalenten bzw. homöopolaren bzw. Atombindung, 2. der heteropolaren bzw. Ionenbindung und 3. der Metallbindung. Diese drei Typen der Bindung sind als Grenzfälle zu betrachten, sie sind überwiegend nur anteilig in den existierenden Bindungen enthalten.

Die molekülbildenden Kräfte werden auch als *Hauptvalenzkräfte* bezeichnet. Sie realisieren Bindungsenergien von ≈ 1 bis 10 eV. Zudem existieren noch *Nebenvalenzkräfte*, die eher für den Zusammenschluß von Molekülen zu größeren Einheiten verantwortlich sind. Die sie charakterisierenden Bindungsenergien liegen bei nur 0,01 bis 0,1 eV.

5.7.2.1. Atombindungen. Diese Molekülbildungsart ist die weitaus häufigste. Sie beschreibt die Bindung zwischen Atomen gleicher Polarität (homöopolare Atome). Beispiele hierfür sind die Moleküle H_2, N_2, O_2, F_2, CH_4. Die Edelgaskonfiguration wird durch gemeinsame Inanspruchnahme von Elektronen erreicht. Es bilden sich zu jeweils zwei Partnern im Molekül gehörige Elektronenpaare aus. Diese bindenden Elektronenpaare sind durch einen Valenzstrich dargestellt. Abb. 5.7.2–1 veranschaulicht die Entstehung der Valenzstrich-Darstellung. In beiden Beispielen wird durch das gemeinsame Elektronenpaar die Edelgaskonfiguration erreicht. Im Falle des Wasserstoffes ist es ausnahmsweise nicht das Oktett, sondern das Helium-Dublett. Atombindungen haben in symmetrischer Anordnung kein permanentes Dipolmoment, da die bindenden Elektronen auf der Verbindungslinie zwischen den Atomen angeordnet sind. Die Vorstellung geht von um die Verbindungsachse zwischen den Atomen kreisenden Elektronen aus. Aus dem Blickwinkel des Orbitalmodelles für Wasserstoff rücken die Kerne zusammen und werden gemeinsam von beiden Elektronen umkreist. Die Anschauung liefert auch hierfür kein permanentes Dipolmoment.

Es kommt auch zur Ausbildung von Doppel- und Mehrfachbindungen. Zwei Sauerstoffatome mit ihren sechs äußeren Elektronen erreichen die Edelgaskonfiguration nur durch zwei gemeinsame Elektronenpaare, was durch die Schreibweise mit zwei Valenzstrichen ausgedrückt wird, O = O. Beim Stickstoff sind es drei Elektronenpaare, und

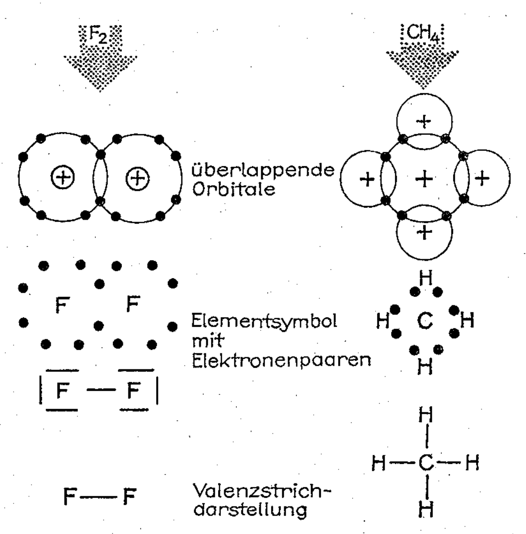


Abb. 5.7.2–1 Valenzbindung (Atom- bzw. homöopolare Bindung)

dementsprechend schreibt man $N \equiv N$. Die höheren Elemente der dritten und vierten Hauptgruppe im Periodensystem der Elemente bilden oft keine Oktett-Bindung aus. Sie bilden eine der Helium-Edelgaskonfiguration analoge äußere Schale mit zwei Elektronen.

Der Versuch, die homöopolare Bindung mit den Mitteln der klassischen Physik zu erklären, liefert keine befriedigenden Ergebnisse. Entsprechend dem Valenzstrich befinden sich die bindenden Elektronenpaare zwischen den Atomen. Sie ziehen die ohne sie positiven Ionen an. Die Rechnungen dazu liefern aber wesentlich zu kleine Werte für die Bindungsenergie, wenn man die bekannten Atomabstände einsetzt. Die quantentheoretische Erklärung für Atombindungen beruht auf Austauschwechselwirkungen. Die klassische Physik hat dazu nichts Analoges. Einfachster Fall hierfür ist das Wasserstoffmolekülion, in dem die zwei Protonen von einem Elektron zusammengehalten werden. Die Vorstellung dazu ist, daß das Elektron die Atomzugehörigkeit in schneller Folge wechselt und aus dem Austausch eine bindende Kraft resultiert. Eine Analogiebetrachtung – wo zwar kein Teilchen, sondern nur Energie ausgetauscht wird – stammt von KOSSEL. Gekoppelte Pendel, die gegeneinander verschiebbar angeordnet sind, bewegen sich im Verlaufe des wechselseitigen Pendelns aufeinander zu. Diese anziehende Kraft realisiert den Übergang in die energetisch günstigste Pendelanordnung.

5.7.2.2. Heteropolare Bindung bzw. Ionenbindungen. Auch hier ist das Bildungsprinzip das Erreichen der Edelgaskonfiguration. Es verbinden sich Atome, die Ionen verschiedener Polarität bilden. Ein Atom gibt die zum Schalenabschluß überzähligen Elektronen ab, ein anderes nimmt sie zum Erreichen des Schalenabschlusses auf. Demgemäß verbinden sich bei zweiatomigen Molekülen Elemente der ersten Hauptgruppe mit denen der siebten, Elemente der zweiten Hauptgruppe mit denen der sechsten und Elemente der dritten mit denen der fünften Hauptgruppe. Abb. 5.7.2–2 zeigt den Vorgang des heteropolaren Zusammenschlusses. Das Kalium hat in der N-Schale nur ein 4s-Elektron. Beim Chlor ist die äußere, die M-Schale, mit zwei 3s- und fünf 3p-Elektronen besetzt. Wechselt das 4s-Elektron des Kaliums in den letzten möglichen 3p-Zustand des Chlors über, so erreichen beide Elemente die Edelgaskonfiguration. Das Molekül ist elektrisch neutral.

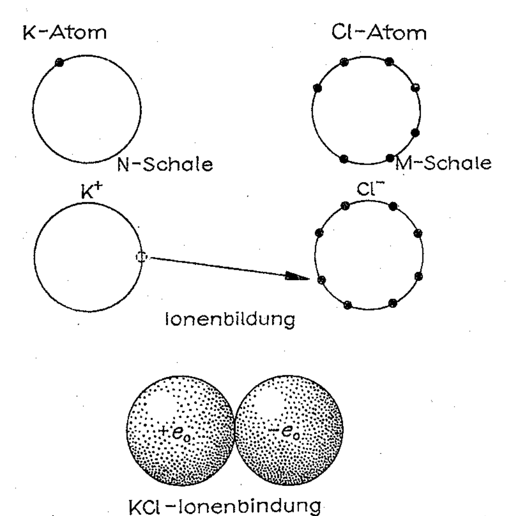


Abb. 5.7.2–2 Entstehung der heteropolaren Bindung bzw. Ionenverbindung

Der Zusammenhalt ist durch die Coulomb-Kraft gegeben. Der Gleichgewichtsabstand stellt sich durch die abstoßenden Kerne ein, wenn die Ionenrümpfe sich etwas durchdringen. Mit den Mitteln der klassischen Physik sind, wie bei der homöopolaren Bindung, die Bindungsenergiewerte nicht richtig zu ermitteln. Die Rechnung erfolgt so, daß die Energiebilanzen der Teilschritte zur Bildung des Moleküls aufgestellt werden. Die Ionisierungsenergie zur Erzeugung des positiven Ions ist aufzubringen. Ein Teil dieses Betrages wird frei bei der Elektronenanlagerung für das negative Ion (Elektronenaffinität). Bei Annäherung der molekülbildenden Ionen wird über die anziehende Coulomb-Kraft ein weiterer Energiebetrag frei, so daß die gesamte Energiebilanz positiv ist, Bindungsenergie frei wird, die Bindung sich ausbildet.

Tab. 5.7.2–3 Elektronegativitäten einiger Elemente

Cs	K	Na	Li	Mg	Al	Si	H	Au	Br	N	O	F
0,7	0,8	0,9	1,0	1,2	1,5	1,8	2,1	2,4	2,8	3,0	3,5	4,0

Die heteropolare Bindung liefert große, permanente Dipolmomente. Das Dipolmoment gestattet es, den heteropolaren Bindungsanteil in gemischten Bindungen quantitativ anzugeben. Sie liegen immer unter den zu erwartenden Werten für sich aneinanderlagernde Ionen. Qualitativ erklärt sich das dadurch, daß die Ionen sich unter dem Einfluß der gegenseitigen Anziehung polarisieren. Die so entstehenden inneren Dipole sind dem bindenden Dipol entgegengerichtet und verringern das resultierende Moment. Eine reine Ionenverbindung existiert nicht. Als Maß für den heteropolaren Bindungsanteil ist die *Elektronegativität* geeignet. Sie ist ein relatives Maß für die Kraft, mit der von einem Atom die Bindungselektronen des anderen Atoms angezogen werden. Die Elektronegativität der Atome ist berechenbar und auch tabelliert. Tab. 5.7.2–3 gibt die Werte für einige Atome an. Je größer die Differenz der Elektronegativitäten zwischen gebundenen Atomen ist, um so größer ist der ionische Bindungsanteil. Der Differenzwert Null beschreibt die reine Atombindung, die Grenze des Überwiegens der Atombindung liegt bei 1,7, und 3,0 kennzeichnet 90 % heteropolaren Bindungsanteil. Fluor hat definitionsgemäß den Wert 4,0.

Aufgrund der richtungsunspezifischen Kraftwirkung zwischen Ionen wird oft zur Unterscheidung gegen die gerichtete Bindung vom homöopolaren Typ, der Ausdruck „Ionenbindung" vermieden und dafür der Begriff *Ionenverbindung* (Ionenbeziehung, Elektrovalenz) verwendet. Die ungerichtete Bindung führt auch dazu, daß sich eher Salze (Ionenverbindungen) als Moleküle bilden. Die bessere Raumfüllung der Salze führt zu einer Erhöhung der Bindung relativ zu NaCl-Ionen. Der Kristall ist energetisch bevorzugt und deshalb überwiegend ausgebildet. Beispielsweise im NaCl hat jedes Na-Ion sechs Cl-Atome als nächste Nachbarn, und analog hat jedes Cl-Atom sechs Na-Ionen als Nachbarn.

5.7.2.3. Metallbindung. Dieser Bindungstyp beruht wie der heteropolare Zusammenschluß auf ungerichteten Kräften und führt genauso eher zur Ausbildung des Festkörpers, als

zur Molekülbildung. Die Moleküle kondensieren zum Kristall, bilden seine Dampfphase.
Die Edelgaskonfiguration wird durch Abgabe von nur Elektronen erreicht. Die Atomrümpfe bilden für sich das Metallgitter, in dem sich die abgegebenen Elektronen delokalisiert, frei beweglich aufhalten. Das „Elektronengas" bewirkt den Zusammenhalt und die Ladungsneutralität. Aufgrund seiner freien Beweglichkeit kann es sogar als Kollektiv Eigenschwingungen um die starren Gitterbausteine ausbilden (Elektronenplasmaresonanz). Herausragende physikalische Eigenschaften der Metallbindung sind die gute elektrische und thermische Leitfähigkeit. Sie sind durch die im Metall frei beweglichen Elektronen anschaulich erklärt. Da die Bindungskraft weder gerichtet, noch absättigbar ist, bevorzugen die Atome die dichtesten Packungsgeometrien. Die Mehrzahl der Metalle bildet die kubisch-dichteste Kugelpackung aus, z. B. Kupfer, Gold und Aluminium. Von etwa gleichguter Raumfüllung ist die hexagonal-dichteste Kugelpackung, z. B. des Magnesiums. Noch ≈ 92 % der Raumfüllung dieser dichtesten Packung erreicht der kubisch-raumzentrierte Gittertyp. Nach seinem Schema kristallisieren z. B. die Elemente Kalium, Barium und Wolfram. Nicht alle Metallatome setzen so viele Elektronen frei, daß sie die Edelgaskonfiguration erreichen. Die Elektronenkonfigurationen der Atomrümpfe sind aber ausreichend kugelsymmetrisch, so daß die Bindungskräfte auch nicht gerichtet sind und die dichteste Kugelpackung gleichermaßen ausgebildet wird. Metallatome mit etwa dem gleichen Atomradius und der gleichen Anzahl freiwerdender Elektronen sind im Gitter gegeneinander austauschbar. Die entsprechenden Atome legieren leicht, wobei jedes Mischungsverhältnis möglich ist und die Atome sich statistisch verteilen. Intermetallische Verbindungen liegen dagegen vor, wenn sich feste Zahlenverhältnisse einstellen (Messing). Hier ist der Bindungscharakter nicht mehr rein metallisch, ein heteropolarer Bindungsanteil bestimmt das Zahlenverhältnis. (Messing bildet verschiedene Phasen aus, z. B. CuZn, Cu_4Zn_8 und $CuZn_3$.)

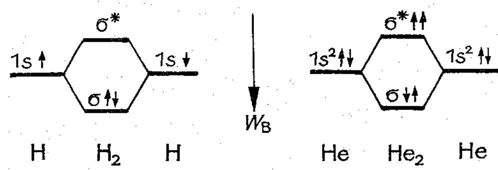
5.7.2.4. Nebenvalenzbindungen. Die bisher besprochenen „echten" Molekülbindungen realisieren Atomabstände von etwa dem Durchmesser der Atome, und es kommt bei der Ausbildung der Edelgaskonfiguration zu einer wesentlichen Umordnung der äußeren Elektronen. Nebenvalenzbindungen bewirken keine so große Annäherung der Atome (nur 0,3 bis 0,5 nm), und die Bindungskräfte betragen lediglich ≈ 0,01 – 0,1 eV. Es handelt sich überwiegend um zwischenmolekulare Kräfte, um Kräfte zwischen valenzmäßig abgesättigten Molekülen. Sie bewirken beispielsweise die Bindung zwischen Edelgasmolekülen bei deren Verflüssigung und sind für Kohäsion und Adhäsion verantwortlich. Auch der Unterschied zu den idealen Gasen in der van der Waalsschen Zustandsgleichung wird von diesen Kräften 2. Ordnung erzeugt. Man spricht auch von van der Waalsschen Kräften und demgemäß von *van-der-Waals-Molekülen.* Diese Moleküle mit nur sehr geringer Bindungsenergie dissoziieren zum Teil schon bei Zimmertemperatur (entsprechend 0,025 eV). Die Bindungsenergie des HCl-Moleküls beträgt ≈ 0,02 eV. Weitere typische Vertreter sind Hg_2 und Cl_2. Van-der-Waals-Kräfte beruhen auf Dipol-Dipol-Wechselwirkung. Drei Arten sind zu unterscheiden. Die Wechselwirkung zwischen zwei permanenten Dipolen, die zwischen einem permanenten und einem induzierten Dipolmoment und die zwischen zwei induzierten Momenten. Charakteristisch für alle drei Arten ist das r^{-6}-proportionale Wechselwirkungspotential.

Wasserstoffbrückenbindungen liegen in der Bindungsenergie zwar unterhalb der chemischen Bindungen, sind mit bis zu 0,5 eV aber größer als die van-der-Waals-Bindungen. Sie bewirken beispielsweise die Festigkeit des Eises und sind intra- wie intermolekular zu finden.
Wasserstoffatome können zwischen stark elektronegativen Elementen bindend wirken, so wie das Elektron zwischen den beiden Protonen im Wasserstoffmolekülion. Stark elektronegative Elemente sind Fluor, Stickstoff und insbesondere Sauerstoff. Diese Elemente ziehen das Elektron des Wasserstoffs stark an. Das Wasserstoffatom ist sehr klein, so daß es zu einer großen Annäherung kommt und sich eine kräftige Bindung ausbildet. Abb. 5.7.2–4 zeigt – durch Schraffur hervorgehoben – die zwei Wasserstoffbrücken der Essigsäure. Die gestrichelt gezeichneten Pfeile deuten an, daß der Valenzstrich zu gleicher Wahrscheinlichkeit beiden Nachbarn gehört. Analog der homöopolaren Bindung sind die Wasserstoffatome der Brücken den jeweils zwei angrenzenden Sauerstoffatomen gleichermaßen zuzuordnen und demgemäß bindend.

Abb. 5.7.2–4 Wasserstoffbrückenbindung am Beispiel des Doppelmoleküls der Essigsäure $[CH_3COOH]_2$

5.7.3. Theoretische Beschreibung der Moleküle

Nach dem allgemeinen Stabilitätsprinzip ist es allen Teilchen und Teilchensystemen eigen, den niedrigst möglichen Zustand einzunehmen. Die Möglichkeit der Atome, im Verbund einen Zustand niederer Energie erreichen zu können, ist Ursache für die Entstehung der chemischen Bindung. Wenn sich ein Molekül unter Energieabgabe an die Umgebung bilden kann, existiert es auch. Umgekehrt wird sich ein Molekül auch nicht bilden, bzw. nicht zeitlich stabil existieren, wenn es mit Energieabgabe an die Umgebung zerfallen kann. Abb. 5.7.3–1 veranschaulicht dies am Beispiel der Energieniveauschemata für Wasserstoff- und Heliumatome bzw. ihrer Moleküle. Für Molekülorbitale wählt man zur Unterscheidung gegen die der Atome für die Bezeichnung der Bahndrehimpulsquantenzahlen griechische Buchstaben. Dementsprechend bezeichnen σ, π, δ Orbitale mit den Drehimpulswerten $\lambda = 0$; 1 bzw. 2. Jedes Niveau in der Abbildung kann von zwei Elektronen unterschiedlicher Spinrichtung (↑ = Spin auf bzw. ↓ = Spin ab) besetzt werden.
Für das H_2-Molekül stehen dafür zwei Niveaus zur Verfügung. Es addieren sich die Niveaus der Atome. Auch bei gleichen Atomen liegen diese Niveaus nicht auf gleicher Höhe im Molekül, die Niveaus spalten auf. Das untere Niveau ist, da es unter dem der Atome liegt, von bindender Natur. Das um etwa den gleichen Betrag über dem der Atome liegende und mit einem Stern gekennzeichnete Niveau ist antibindend. Die beiden einzelnen Elektronen der Atome des Wasserstoffs gehen in das bindende Molekülniveau über. Im Falle des Heliums passen von den vier Elektronen nur zwei in das bindende Molekülniveau, die anderen beiden nehmen den antibindenden Zustand ein. Das ist immer noch die energetisch günstigste Möglichkeit. (Der



Energieniveauschema der Atom- und Molekülorbitale von Wasserstoff und Helium.

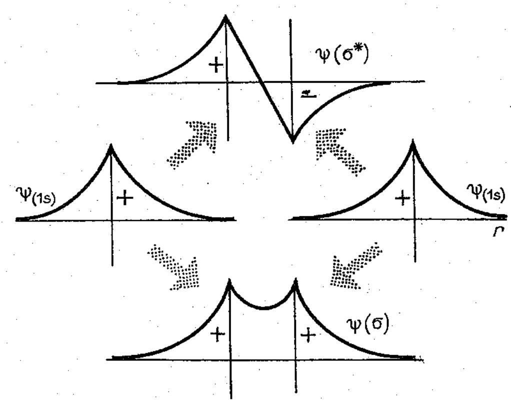


Abb. 5.7.3–1 Skizze der 1s-Wellenfunktionen und ihre Addition zu denen des Moleküles

nächstmögliche, wenn auch als bindend bezeichnete Zustand würde in der Abbildung noch weiter oben einzuzeichnen sein.) Dabei entsteht aber insgesamt kein Gewinn an Energie, so daß die Molekülbildung nicht zustande kommt.

Schon aus dieser einfachen Darstellung ist nun auch verständlich, warum dennoch das Heliummolekülion He_2 existiert. Das eine Elektron, welches fehlt, fehlt im antibindenden Zustand, wodurch sich die Gesamtbilanz der Bindung verbessert und das Molekülion relativ zu einem Atom und einem Ion günstiger liegt. Das bisher Gesagte erhellt die Notwendigkeit, die Energie der Molekülzustände, d. h. die Energieeigenwerte, durch Lösen der Schrödinger-Gleichung zu ermitteln. Im Folgenden sei nur kurz angedeutet, daß auch die Kenntnis der Wellenfunktionen selbst, zum Verständnis der chemischen Bindung unumgänglich ist.

Abb. 5.7.3–1 enthält im unteren Teil skizziert die Psi-Funktionen der Zustände zu den Energieniveaus der Wasserstoffatome und des Moleküls, die im oberen Teil des Bildes gegeben sind. Die Aufenthaltswahrscheinlichkeiten der Elektronen im Atom bzw. Molekül sind durch das Quadrat der abgebildeten Funktionen gegeben. Die Wellenfunktionen der Molekülorbitale entstehen durch Addition der Atomwellenfunktionen. Je nachdem, ob die Vorzeichen der beiden Atomwellenfunktionen gleich oder verschieden sind, entstehen bindende (σ) oder antibindende (σ^*) Molekülorbitale. Die Aufenthaltswahrscheinlichkeiten der Elektronen in einem Molekül der beiden Atome a und b bilden sich somit aus $\Psi^2(\sigma^*) \sim [\Psi_a(1\,s) - \Psi_b(1\,s)]^2$ bzw. $\Psi^2(\sigma) \sim [\Psi_a(1\,s) + \Psi_b(1\,s)]^2$. Man erkennt, daß sich bei der Addition von Funktionen gleichen Vorzeichens ein Gebiet erhöhter Aufenthaltswahrscheinlichkeit zwischen den Atomen bildet. Die Elektronen finden dort ein energetisch günstiges Gebiet, was als eine Begründung des Valenzstriches für zwei bindende Elektronen gewertet werden kann. In der Praxis bedeutet dies, daß sich ein Wasserstoffmolekül nur bilden kann, wenn zwei Wasserstoffatome mit antiparalleler Spineinstellung aufeinander treffen. Die Addition von Funktionen verschiedenen Vorzeichens führt zu einem Nulldurchgang für die Molekül-Psi-Funktion und dadurch zu einem Punkt mit der Aufenthaltswahrscheinlichkeit Null zwischen den Atomen. Die Elektronen sind überwiegend in antibindenden Räumen. Es kommt nicht zur Molekülbildung. Die Entscheidung, ob Orbitale bindend oder antibindend sind, erfordert die Lösung der Schrödinger-Gleichung des Problems, die Ermittelung der Eigenfunktionen des Moleküls. Damit ergibt sich auch der Grad der Überlappung der Atomwellenfunktionen, ein Maß für die Stärke der Bindung.

5.7.3.1. Das Wasserstoffmolekül. Der einfachste Fall des aus zwei identischen Atomen mit nur je einem Elektron bestehenden Moleküls ist das Modellbeispiel für die Behandlung der zweiatomigen Moleküle allgemein. Dabei zeigen sich schon die Erweiterungsmöglichkeiten auf kompliziertere Moleküle. Modellmäßig ist das Beispiel auch für die Art des Herangehens bei der wellenmechanischen Beschreibung der chemischen Bindung. Die Lösung der Eigenwertgleichung $\boldsymbol{H}\Psi_n = W_n\,\Psi_n$ (Hamilton-Operator $\boldsymbol{H}$ angewendet auf die Zustandsfunktionen des Gesamtsystems Ψ_n ist gleich den Energieeigenwerten W_n multipliziert mit Ψ_n) ist nur über Näherungen möglich. Eine erste, relativ unkritische Näherung geht davon aus, daß die leichten, sich schnell bewegenden Elektronen unabhängig von den dazu trägen Kernbewegungen behandelt werden können (adiabatische Näherung). Zur Berechnung etwa von Schwingungen und Vibrationen des Moleküls, also dort, wo die Kernbewegungen entscheidend sind, werden Mittelwerte über die dann bekannten Elektronenzustände eingesetzt. Von den Elektronenfrequenzen her betrachtet, ruht der Kern, von Kernbewegungen aus gesehen, sind die Elektronen so schnell, daß nur Zeitmittelwerte ihrer Bewegung den Kern beeinflussen.

Zur Berechnung der Elektronenzustände wird folgender Hamilton-Operator angesetzt (vgl. 5.6.4.2.), dessen einzelne Glieder mit Abb. 5.7.3–2 verständlich werden, $\boldsymbol{H} = -\hbar^2/2\,m\,(\Delta_1 + \Delta_2) - e^2/4\,\pi\,\varepsilon_0\,(1/r_{a1} + 1/r_{b2} + 1/r_{a2} + 1/r_{b1} - 1/R - 1/r_{12})$. Die Quadrate der Impulsoperatoren, nämlich Δ_1 bzw. Δ_2 entsprechen denen der einzelnen Atome. Die Indizes 1 bzw. 2 weisen darauf hin, daß sich Δ_1 nur auf das Elektron 1, bzw. Δ_2 auf das Elektron 2 bezieht. Die Wechselwirkung zwischen den Atomen wird durch die vier letzten Terme der potentiellen Energie bestimmt. Zwei Glieder berücksichtigen die anziehenden Kräfte zwischen dem Kern a und dem Elektron zum Kern b ($1/r_{a2}$) und zwischen dem Kern b und dem Elektron zum Kern a ($1/r_{b1}$), und die anderen beiden

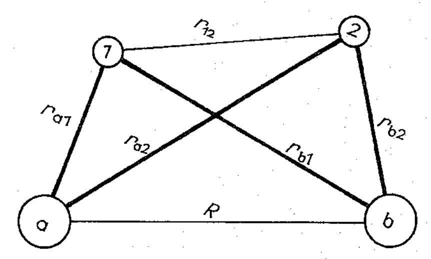


Abb. 5.7.3–2 Bezeichnung der Abstände im Wasserstoffmolekül mit den Kernen a, b und den Elektronen 1, 2

Glieder realisieren die Kernabstoßung ($-1/R$) bzw. die Abstoßung der Elektronen ($-1/r_{12}$).
Nach der Art, die Zustandsfunktionen Ψ_n des Moleküls zu bilden, unterscheidet man im wesentlichen 2 verschiedene Vorgehensweisen, die günstigste Näherung anzusetzen. Die Theorie der *Valenzstrukturen* – *VB-Methode* (von engl. *v*alence *b*ond) – bildet die Wellenfunktionen aus einer Linearkombination von Produkten der Atomwellenfunktion. Die durch die Bindung eintretende Umordnung der Elektronenzustände wird nachträglich durch eine Störungsrechnung berücksichtigt. Bedingung ist, daß die Störung hinreichend gering bleibt. Ein Verfahren also, das für große Atomabstände von Vorteil ist. In der Theorie der *Molekülorbitale* – *MO-Methode* (von engl. *m*olecular *o*rbital) – bildet man erst den Rumpf bzw. Kern des Moleküls aus den Rümpfen bzw. Kernen der Atome und ordnet diesem die zur Verfügung stehenden Elektronen zu. Bei beiden Ansätzen ist so zu variieren, daß Wellenfunktionen gebildet werden, die minimale Energiezustände liefern.
Der Ansatz entsprechend der MO-Theorie zur Bestimmung der Wellenfunktion des Wasserstoffmoleküls ergibt sich mit $\Psi \sim (\Psi_{a1} + \Psi_{b1})(\Psi_{a2} + \Psi_{b2})$. Die erste Klammer beschreibt das Elektron 1 im Feld der Kerne a und b und die zweite Klammer das Elektron 2 im gleichen Molekülkernfeld. Die Klammer ausmultipliziert und mit Koeffizienten versehen, ergibt $\Psi = A\Psi_{a1}\Psi_{b2} + B\Psi_{b1}\Psi_{a2} + C\Psi_{a1}\Psi_{a2} + D\Psi_{b1}\Psi_{b2}$. In diesem allgemeinen Ansatz stehen die ersten beiden Summanden für den homöopolaren Bindungsanteil und die letzten beiden Summanden für den heteropolaren Bindungsanteil. Wählt man die Koeffizienten $C = D = 0$, so existieren nur noch Produktwellenfunktionen, bei denen jeweils nur ein Elektron zu einem Kern gehört, berücksichtigt der Ansatz nur unpolare Glieder. Im Falle $A = B = 0$ existieren nur Wellenfunktionen, bei denen beide Elektronen nur entweder zum Kern a oder zum Kern b gehören, es wird die reine Ionenbindung beschrieben. Beides sind Grenzfälle des allgemeinen Ansatzes, des Regelfalles, bei dem alle Koeffizienten $\neq 0$ sind.
LCAO-Methode. Entsprechend dem Vorgehen nach der VB-Theorie, kann man in nullter Näherung für die Moleküleigenfunktion eine Linearkombination nicht miteinander wechselwirkender Atome ansetzen und die Wechselwirkung anschließend als Störung einführen. Das Verfahren der Linearkombination von Atomeigenfunktionen wird nach dem englischen Ausdruck „*l*inear *c*ombination of *a*tomic *o*rbitals" als ICAO-Methode bezeichnet. Es ergeben sich zwei Möglichkeiten, die Ausgangswellenfunktionen linear zu kombinieren: $\Psi_+ = A\Psi_{a1}\Psi_{b2} + B\Psi_{a2}\Psi_{b1}$ und $\Psi_- = A\Psi_{a1}\Psi_{b2} - B\Psi_{a2}\Psi_{b1}$. Wegen der Ununterscheidbarkeit der Teilchen wird $A = B$. Da die Gesamtwellenfunktion antisymmetrisch sein muß (Pauli-Prinzip), gehört zu der in den Ortskoordinaten symmetrischen Form Ψ_+ eine antisymmetrische Spinausrichtung ($\uparrow\downarrow$) und zu Ψ_-, der bei Vertauschung der Ortskoordinaten antisymmetrischen Wellenfunktion, gleichgerichtete Spins der beiden Atome ($\uparrow\uparrow$). Für den Fall der antiparallelen Spinausrichtung ($\downarrow\uparrow$) ist der Gesamtspin $S = 0$, und für parallele Spinausrichtung ($\uparrow\uparrow$) ist der Gesamtspin $S = 1$.
In nullter Näherung sind die beiden Atome des Moleküls noch wechselwirkungsfrei, und der Hamilton-Operator ist von der einfachen Form $H^0 = -h^2/2\,m(\Delta_1 + \Delta_2) - e^2/(4\pi\varepsilon_0)(1/r_{a1} + 1/r_{b2})$. Der Rest des Operators zu dem schon angegebenen vollständigen Hamilton-Operator $\boldsymbol{H}$, $\boldsymbol{H}' = -e^2/(4\pi\varepsilon_0)(1/r_{a2} + 1/r_{b1} - 1/R - 1/r_{ab})$, ist als eine die Wellenfunktionen und Energiewerte nur schwach verändernde Störung anzusetzen. Die Durchrechnung des Problems ergibt die Energieeigenwerte $W_\pm = 2\,W_H + (W_C \pm W_A)/(1 \pm S^2)$. W_H bezeichnet die Energie des isolierten Wasserstoffatoms.

$$W_C = \int \Psi_{a1}^* \Psi_{b2}^* \boldsymbol{H}' \Psi_{a1} \Psi_{b2}\, dV \text{ (Coulomb-Integral)}$$

$$W_A = \int \Psi_{a1}^* \Psi_{b2}^* \boldsymbol{H}' \Psi_{a2} \Psi_{b1}\, dV \text{ (Austauschintegral)}$$

$$S^2 = \int \Psi_{a1}^* \Psi_{b2}^* \Psi_{a2} \Psi_{b1}\, dV \text{ (Überlappungsintegral)}$$

Das Coulomb-Integral liefert den auch für klassische Rechnungen zu erwartenden Coulomb-Anteil der Wechselwirkungsenergie, den Anteil aus der Kern-Kern- und Elektron-Elektron-Abstoßung und den der Anziehung des Elektrons 1 vom Kern b und des Elektrons 2 vom Kern a. Austausch- und Überlappungsintegral sind quantenphysikalischer Natur. Erst durch ihre Existenz spalten die Energieeigenwerte auf. W_+ kennzeichnet dabei die Energie bei verschiedenen Spinrichtungen der beiden Elektronen ($\uparrow\downarrow$) und W_- die gleichgerichteter Spins ($\uparrow\uparrow$). Das Austauschintegral ist durch den gleichen Hamilton-Operator der Wechselwirkung zwischen den Atomen ähnlich dem Coulomb-Integral, nur zeigt die Bezeichnung der Wellenfunktionen an, daß ein Energieanteil zur Gesamtenergie durch Elektronenaustausch zu berücksichtigen ist. Es existiert eine Austauschwechselwirkung, die letztendlich durch die Ununterscheidbarkeit der Elektronen hervorgerufen wird. Da man nicht entscheiden kann, welches Elektron zu welchem Kern gehört, sind beide Möglichkeiten a1, b2 und a2, b1 zuzulassen. Dieser Elektronenaustausch beeinflußt die Gesamtenergie. Das Überlappungsintegral S^2 ist dimensionslos. Die Größe des Integrals ist von der Stärke der räumlichen Überlappung der Elektronenbahnen abhängig. W_A und S^2 sind proportional und verschiedenen Vorzeichens. Abb. 5.7.3–3 zeigt den Verlauf der Potentialkurven $W_+(R)$ bzw. $W_-(R)$. Dargestellt sind die experimentell ermittelten Werte. Die Abbildung zeigt an, daß, wenn zwei Wasserstoffatome mit

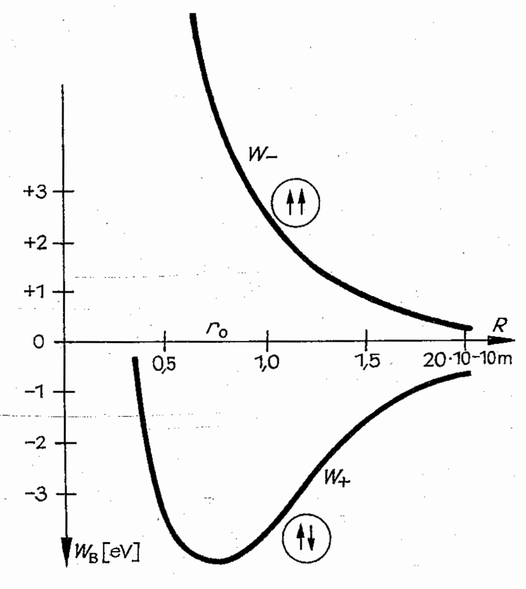


Abb. 5.7.3–3 Bindungsenergien des H_2-Moleküles als Funktion des Atomabstandes R

zueinander paralleler Spinausrichtung (↑↑) sich einander nähern, zur Verringerung des Abstandes R Energie aufgebracht werden muß. Die Atome gehen miteinander keine Bindung ein, sie stoßen sich ab. Ein anderer Potentialverlauf kennzeichnet die Begegnung zweier Wasserstoffatome mit zueinander antiparalleler Spineinstellung (↑↓). Im Abstand von etwa $R = 0{,}74 \cdot 10^{-10}$ m sind die Atome mit $\approx 4{,}5$ eV Bindungsenergie zu einem stabilen Molekül verbunden. Im Tal der Potentialkurve befinden sich die Atome in einem Bindungszustand mit dem Gleichgewichtsabstand r_0. Bei Verringerung des Abstandes verringert sich die Bindung durch die Kernabstoßung, bei Vergrößerung des Abstandes verringert sich die Bindung durch die geringere Überlappung der Wellenfunktionen, durch einen verringerten Betrag der Austauschwechselwirkungsenergie. Der hier angeführte sehr einfache Rechengang liefert nur eine Bindungsenergie von $W_+ = 2{,}9$ eV, und der Gleichgewichtsabstand ergibt sich zu $\approx 0{,}88 \cdot 10^{-10}$ m. Es wird aber schon ein bindender Zustand erreicht, also die Bindung des H_2-Moleküls begründet. Durch Berücksichtigung polarer Anteile von $\approx 5\,\%$ ($C \neq 0$, $D \neq 0$) kann der Fehler etwas verringert werden. Um die Genauigkeit experimentell ermittelter Werte der Bindungsenergie zu erreichen, ist die Näherungsmethode in der vorgestellten einfachen Form nicht ausreichend. Es genügt nicht, nur die unveränderten Atomwellenfunktionen zu verwenden. Die Deformation der Elektronenwolken unter dem Einfluß des Partneratoms ist zu berücksichtigen. Allgemein ist zu bemerken, daß zwar dieser oder jener theoretische Ansatz von Fall zu Fall bevorzugt wird, rechentechnische Vorteile bieten kann, aber auf die Genauigkeit der Ergebnisse bezogen, die Unterschiede beispielsweise zwischen der MO- und der VB-Theorie nicht bedeutend sind. Weiterentwickelte Theorien verknüpfen die Ansätze der MO- und VB-Theorien.

5.7.3.2. Form und Lage einfacher Molekülorbitale. Nach der Formel für die Energien der Molekülorbitale des Wasserstoffmoleküls läßt sich verallgemeinernd ein „Prinzip der maximalen Überlappung" formulieren. Es überlagern sich immer die Eigenfunktionen der ein Elektronenpaar bilden-

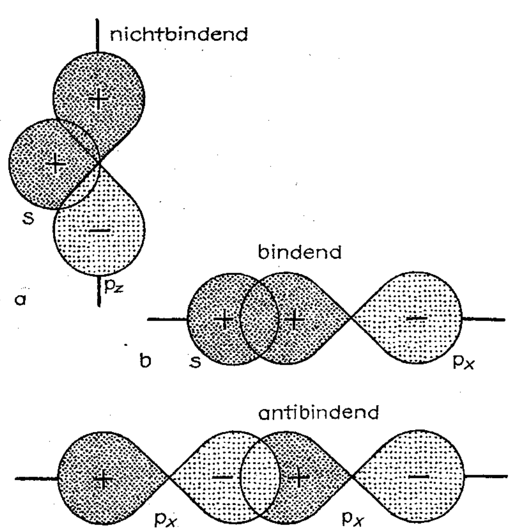


Abb. 5.7.3–4 Die verschiedenen Bindungstypen je nach Art der Überlappung von Atomorbitalen

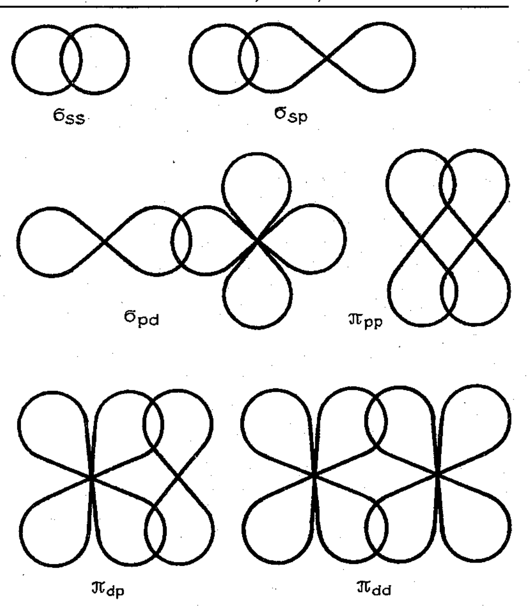


Abb. 5.7.3–5 Verschiedene Kombinationen von Atomorbitalen zu σ- und π-Bindungen

den Elektronen zweier Atome. Je größer die Überlappung der Atomorbitale, um so fester die Bindung. Die maximale Überlappung wird erreicht, wenn die Orbitale ähnliche Energien haben und von gleicher Symmetrie bezüglich der Bindungsachse des Moleküls sind. Ähnliche Energien, d. h. die gleiche Hauptquantenzahl, sind als Interferenzbedingung für eine ungefähr gleiche Frequenz zu fordern. Gleiche Symmetrie kennzeichnet das gleiche Amplitudenvorzeichen, bei verschiedenen Vorzeichen werden antibindende Orbitale ausgebildet. Die größtmögliche Überlappung ergibt sich auf der Symmetrieachse der Orbitale. In Abb. 5.7.3–4 zeigt *a* ein Beispiel verschiedener Symmetrie und *b* einen bindenden und einen antibindenden Zustand. Die Koordinaten sind wie im folgenden so gewählt, daß die Bindungsachse immer in x-Richtung liegt. Im Beispiel *a* heben sich die Überlappungsanteile bindender und antibindender Orbitale in der Summe auf, man spricht von einem nichtbindenden Zustand. Es wird zwischen Einfach- und Mehrfachbindungen unterschieden, je nach der Zahl der sich überlappenden Orbitale. Je eine Bindung wird durch ein gemeinsames Elektronenpaar realisiert. Die Anzahl der Bindungen heißt *Bindungsgrad.*

Molekülzustände sind nach der Symmetrie ihrer Eigenfunktionen bezüglich der Verbindungsachse bezeichnet. σ-Zustände sind rotationssymmetrisch in bezug auf die Verbindungslinie. Sie werden bei Überlappung zweier s-Zustände bzw. von s- und p_x- oder von zwei p_x-Orbitalen gebildet. Der Buchstabe σ zeigt an, daß der Bahndrehimpuls des Molekülorbitals Null ist. Auch die Verbindung s-p_x hat in Richtung der Molekülachse keine Drehimpulskomponente. p_y-p_y- bzw. p_z-p_z-Überlappungen bilden π-Bindungen ($\lambda = 1\,\hbar$). Abb. 5.7.3–5 zeigt Beispiele für σ- und π-Bindungen und eine zugehörige Bezeichnung. Die σ-Bindung ist die stärkere von beiden Arten. Aufgrund der bei σ-Bindungen rotationssymmetrischen Elektronenverteilung zwischen den gebundenen Rümpfen auf der Bindungs-

achse, können sich die Atome frei gegeneinander drehen. Π-Bindungen sind zwar schwächer, aber steifer gegen Verdrehungen, bei ihnen können sich höchstens Torsionsschwingungen ausbilden.
Wie von den Atomzuständen her gewohnt (vgl. z. B. Abb. 5.6.3–2), lassen sich auch die Molekülorbitale in einem Energieniveauschema anordnen. In Abb. 5.7.3–6 sind Zustände für Moleküle aus zwei gleichen Atomen mit von oben nach unten ansteigender Bindungsenergie dargestellt. Der Ausschnitt zeigt Niveaus der Hauptquantenzahl $n = 2$. Die Reihenfolge ist zumindest für die Elemente der 2. Periode so gegeben. Jeder der Zustände ist mit zwei Elektronen unterschiedlicher Spinorientierung besetzbar. Die darüber hinaus entarteten Zustände sind nebeneinander gezeichnet. Die energetische Reihenfolge ist in diesen Fällen $\sigma_{1s} > \sigma_{1s}^* > \sigma_{2s} > \sigma_{2s}^* > \sigma_{2p_x} > \pi_{2p_y} = \pi_{2p_z} > \pi_{2p_y}^* = \pi_{2p_z}^* > \sigma_{2p_x}$. Wie üblich, sind die antibindenden Zustände durch einen Stern gekennzeichnet.
Gewinkelte Valenzen. Die keulenartigen Elektronendichteverteilungen der p- und d-Elektronen im Atom sind Grundlage des Verständnisses vieler Molekülformen, da sich die Bindungen, d. h. die Überlappungen der Atomorbitale, in den Bereichen größter Elektronendichten ausbilden. Bekanntestes Beispiel ist das Wassermolekül. Von den drei zueinander orthogonalen Symmetrierichtungen der 2 p-Zustandsfunktionen des Sauerstoffs, sind zwei mit einem nicht gepaarten Elektron besetzt. An den Keulen ihrer Dichteverteilungen kann sich je ein Wasserstoffatom bindend anlagern. Andere gleichartige Beispiele bilden die Moleküle H_2S und H_2Se, sowie auch die Molekülionen NH_2^- und FH_2^+. Abb. 5.7.3–7 zeigt die zueinander senkrechten 2p-Orbitale des Sauerstoffes und die überlappenden beiden s-Orbitale des Wasserstoffes. Real wird ein Bindungswinkel etwas größer als 104° gemessen, beim NH_3 sind es $\approx 107°$. Beim H_2S liegen die Schenkel nur um 2° und beim H_2Se nur um 1° über dem erwarteten Wert von 90°. Die bisherigen sehr einfachen Betrachtungen unter Verwendung auch im Molekül noch unveränderter Atomorbitale sind zu verfeinern. Das gelingt unter Berücksichtigung von ionischen Bindungsanteilen, auch sind noch die 2s-Elektronen im Sauerstoff zu einem kleinen Prozentsatz an der Bindung mitbeteiligt.

5.7.3.3. Hybridisierungen. Mitunter ist es zur Ausbildung einer starken Bindung energetisch günstiger, diese über einen angeregten Zustand zu erreichen. Für ein Atom ist zwar Energie aufzubringen, um in den angeregten Zustand überzugehen, wenn dadurch aber wesentlich größere Überlappungen mit den Bindungspartnern möglich werden, kann die Bindungsenergie in der Gesamtbilanz größer sein.
Kohlenstoff hat im Grundzustand zwei ungepaarte Elektronen in p-Orbitalen, ist aber gerade durch seine Vierwertigkeit das Grundelement der organischen Chemie. Die Promotionsenergie zum Erreichen eines bindungsgünstigen angeregten vierwertigen Zustandes beträgt 4,2 eV. Hierbei geht ein 2s-Elektron in den dritten freien ungepaarten p-Zustand über, auch das zurückbleibende andere 2s-Elektron ist deshalb ungepaart und bindungsfähig. So ist die Vierwertigkeit des Kohlenstoffes erst einmal erklärt. Nun wäre es verständlich, wenn das freie 2s-Elektron des Kohlenstoffes eine von den drei p-Elektronen verschiedene Bindungsstärke realisieren würde und andere Bindungslängen dort zu finden wären. Dem entgegen ist dem Experiment zu entnehmen, daß im CH_4-Molekül die vier Wasserstoffatome symmetrisch, mit gleicher Bindungsenergie und gleichen

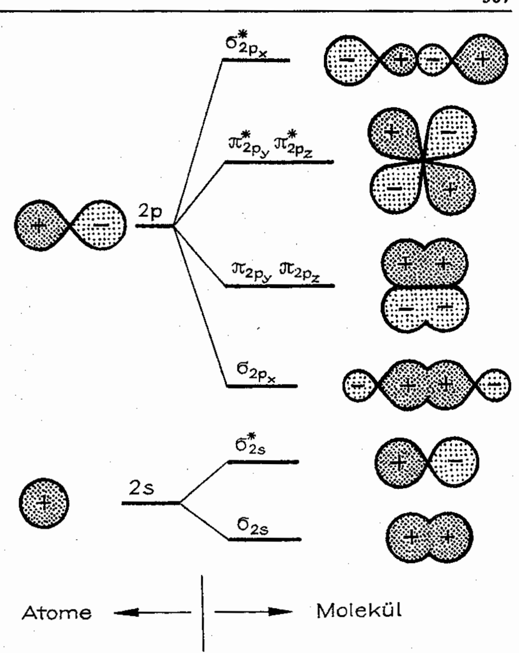


Abb. 5.7.3–6 Energetische Lage und räumliche Verteilung von Atom- und Molekülorbitalen

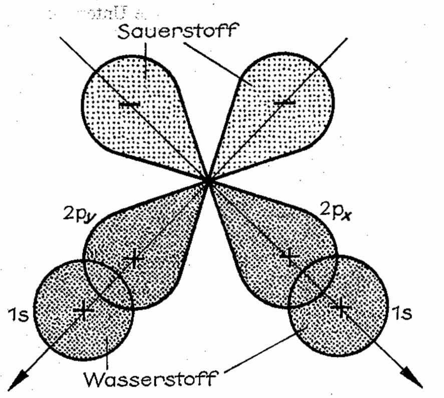


Abb. 5.7.3–7 Bindungsverhältnisse im H_2O-Molekül

Bindungslängen angeordnet sind. Deshalb muß man davon ausgehen, daß die vier Kohlenstofforbitale im CH_4-Molekül gleich sind. Wenn aus verschiedenen Orbitalen im Molekül gleiche entstehen, so spricht man von Hybridorbitalen und nennt den Vorgang *Hybridisierung*. Hybridorbitale haben eine besonders große Ausdehnung in Bindungsrichtung und bewirken damit maximale Überlappungen mit den Partnern im Molekül. Hybridorbitale entstehen in der theoretischen Beschreibung durch Linearkombination der beteiligten Zustände. Die vier Hybridorbitale des Kohlenstoffes kombinieren sich in folgender Form:

$$\Psi_1 = 1/2\,(\Psi_{2s} + \Psi_{2p_x} + \Psi_{2p_y} + \Psi_{2p_z})$$
$$\Psi_2 = 1/2\,(\Psi_{2s} + \Psi_{2p_x} - \Psi_{2p_y} - \Psi_{2p_z})$$
$$\Psi_3 = 1/2\,(\Psi_{2s} - \Psi_{2p_x} + \Psi_{2p_y} - \Psi_{2p_z})$$
$$\Psi_4 = 1/2\,(\Psi_{2s} - \Psi_{2p_x} - \Psi_{2p_y} + \Psi_{2p_z})$$

Diese Veränderungen werden sp^3-Hybridisierung genannt. An ihr sind ein s- und drei p-Elektronen beteiligt. Jedes Hybridorbital hat zu 25 % s- und zu 75 % p-Charakter. Die sp^3-Hybridorbitale des Kohlenstoffes liegen energetisch noch unter dem angeregten Zustand, in dem ein s-Elektron in den dritten freien, ungepaarten p-Zustand übergeht. Die sp^3-Hybridorbitale sind zueinander tetraedrisch angeordnet, diese Form ist z. B. beim Methanmolekül (CH_4) ausgebildet. Die vier freien Kohlenstoffvalenzen binden dabei je ein Wasserstoffatom. Für andere Beispiele realisiert sich eine sp^2-Hybridisierung mit drei beteiligten Elektronen, eins im s- und zwei im p-Zustand. Auch die lineare Form der sp-Hybridisierung von Kohlenstofforbitalen existiert. Abb. 5.7.3–8 zeigt einige Hybridorbitale. Aus der Symmetrie der Formen ergeben sich die Winkel zwischen den Bindungsrichtungen.

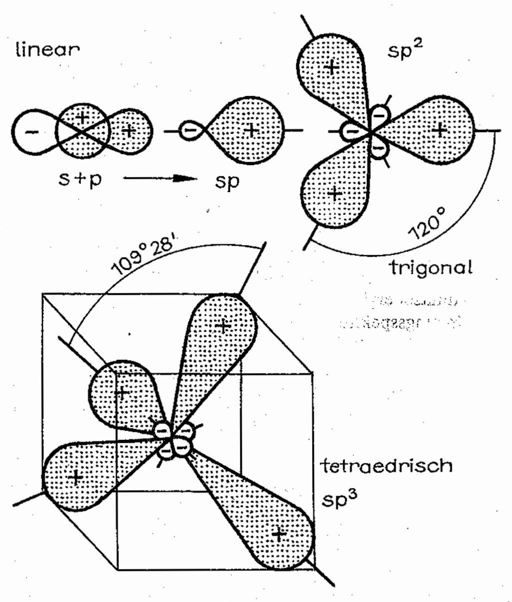


Abb. 5.7.3–8 Einfache Hybridorbitale und ihre Bindungsrichtungen

5.8. Molekülspektroskopie

5.8.1. Anregungsformen und Informationsgehalt der Molekülspektroskopie

Moleküle können sehr sensible Gebilde darstellen und schon bei leichter Wärmezufuhr dissoziieren, zerfallen. Bevor sie dissoziieren, durchlaufen sie verschiedene Anregungsstufen. Diese Anregungen benötigen eine noch geringere Energiezufuhr. Auch sehr stabile Molekülformen sind bekannt, die noch bei Temperaturen von einigen tausend Kelvin existieren. In Mehrheit wird in der Molekülspektroskopie eine feinere Energieabstufung der Anregungen vorhanden sein als bei Atomen, und somit wird in der Regel auch eine höhere Energieauflösung zur Untersuchung der Spektren benötigt. Die einfachsten Moleküle sind zweiatomig, ihre Anregungsformen sind in ihrer Einfachheit Modellfälle zur Einführung in die allgemeine Spektroskopie der Moleküle. Je komplizierter die Moleküle aufgebaut sind, um so vielfältiger sind auch die Anregungsmöglichkeiten. Insbesondere zur Molekülspektroskopie wird der Wellenlängenbereich vom ultravioletten über das sichtbare Gebiet und den Bereich selbst des fernen Infraroten hinaus bis in das Gebiet der Hochfrequenzspektroskopie analysiert. Die Hochfrequenzspektroskopie ist eine vorzügliche Möglichkeit, Molekülstrukturen zu bestimmen. Ihr Auflösungsvermögen von $\nu/\Delta\nu \sim 10^7$ bietet beispielsweise die Möglichkeit, im Bereich zwischen 3 und 300 GHz (10^{-1} bis $10^{-3}\,m^{-1}$) einige 10^4 Spektrallinien noch zu trennen. In der Spektroskopie ist anstelle der Frequenz ν und auch als Energiemaß oft die Wellenzahl $\bar{\nu} = 1/\lambda = \nu/c$ und zudem noch in cm^{-1} angegeben. Die Umrechnung ergibt $1\,eV = 8\,067\,cm^{-1} = 8{,}067 \cdot 10^5\,m^{-1}$.

Alle Anregungszustände der Moleküle sind diskret, sie bestehen aus einzelnen Linien. Wie bei den Atomen kennzeichnet dies die Quantelung der Zustände. Allgemein sind bei einfachen Molekülen klar drei Arten von Spektren zu unterscheiden. Mit dem geringsten Energieaufwand anzuregen, im fernen Ultrarot gelegen, findet man Serien von Linien mit zueinander konstanten Abständen. Sie belegen Anregungen von Rotationszuständen der Moleküle. Mit um eine Größenordnung größeren Abständen erscheinen dann Serien, die Schwingungszuständen zuzuordnen sind. Das gleichzeitige Auftreten beider Arten von Linien führt zur Ausbildung von Rotationsschwingungsspektren der Moleküle. Reine Schwingungsspektren ohne Rotationsterme sind i. allg. Fall nicht zu beobachten, da zu jedem Schwingungszustand Rotationen möglich sind und bei der Schwingungsanregung durch ihren wesentlich geringeren Energiebedarf auch mitangeregt werden. Die größten Abstände aufeinander folgender Linien einer dritten Serienart entstehen durch die Anregung einzelner Molekülelektronen. Das Anheben in energetisch höhere Orbitale (mit geringerer Bindungsenergie) ist analog der Anregung von Atomelektronen.

Die Vielfalt der Rotationsmöglichkeiten, diverse Schwingungsfreiheitsgrade und die große Mannigfaltigkeit der Molekülorbitale führen zu einer verwirrenden Zahl von Linien in Molekülspektren. Durch sie erhält man sehr viele Informationen über die Moleküle. Aus den Rotationsspektren kann man die Trägheitsmomente in bezug auf alle wirkenden Rotationsachsen ermitteln und in einfachen Fällen Kernabstände berechnen. Schwingungsspektren enthalten Informationen über die Kraftkonstanten der Bindung, Dissoziation und auch die Molekülbildung sind in Feinheiten untersuchbar.

Da jedes Molekül ein nur ihm eigenes Gesamtbild der Anregungslinien aufweist, sind prinzipiell alle Moleküle durch spektroskopische Analysen identifizierbar. Selbst ohne Wissen über die Gesetzmäßigkeiten der Spektren kann die Identifizierung analog der Fingerabdruckmethode in der Kriminalistik erfolgen. Man müßte allerdings vorher alle Moleküle mit anderen Methoden analysiert und in Katalogen erfaßt haben. In Kenntnis der Gesetzmäßigkeiten gelingt die Identifizierung auch unbekannter Moleküle. Selbst einzelne Atome kompliziertester Moleküle können durch ihre Bindung im Molekül schwingen und über diese Schwingungen erkannt werden. Auch die Bindungsart kann bestimmt werden. Die dafür prädestinierten Wellenlängenbereiche liegen eher im Infraroten oder bei noch längeren Wellenlängen, im Bereich der Mikrowellen, als im sichtbaren oder ultravioletten Spektralgebiet. Aus den Absorptionsintensitäten gut bekannter Strukturen sind quantita-

tive Analysen möglich, können Mengenanteile beispielsweise in Gasgemischen ermittelt werden.
Es wird in Emission und Absorption spektroskopiert. Die zu untersuchenden Moleküle können beispielsweise in Gasentladungen durch Elektronenstoß angeregt werden, wodurch die Spektren in Emission erscheinen. Für die Absorptionsspektroskopie verwendet man Lichtquellen mit kontinuierlicher Spektralverteilung, beispielsweise Wasserstoffgasentladungslampen. Bequem einsetzbar sind Hochfrequenz- und Mikrowellensender.
Das hier unter Spektroskopie Dargestellte kann nur ein Ausschnitt der Grundlagen sein. Heutzutage sind Molekülreaktionen sogar in ihrem zeitlichen Verlauf beobachtbar (zeitauflösende Spektroskopie). Die dazu eingesetzten Ultrakurzzeitspektrometer arbeiten schon im Subpikosekundenbereich. Damit sind Details der Reaktionsabläufe während der typischen Relaxationszeiten von Schwingungen und Rotationen der Moleküle von $\gtrsim 10^{-10}$ s untersuchbar geworden. Zur Anregung der Moleküle verwendet man Laser, die im Bereich von Pikosekunden pulsbar sind. Auf der Nachweisseite sind schnell arbeitende Fotodioden und Sekundärelektronenvervielfacher eingesetzt. Auch gelingt es, mit Bildwandlerkameras bei elektronenoptischer Bildablenkung (Streakkameras) Abbildungen in zeitlicher Auflösung von 10^{-13} s herzustellen.

5.8.2. Grundarten der Spektren

5.8.2.1. Rotationsspektren. Ein zweiatomiges starres Molekül entspricht dem Hantelmodell. Die Rotation erfolgt um eine Achse senkrecht zur Verbindungslinie der Kerne. Die Rotationsachse geht durch den Schwerpunkt beider Massen. Entsprechend der Quantelung atomarer Drehimpulse ergeben sich die Energiezustände des Rotators zu $W_{rot} = \hbar^2/(2\,\Theta_x)\,J\,(J+1)$ *(Rotationsenergie).*
Mit der Drehimpulsquantenzahl $J = 0, 1, 2, 3, \ldots$ und dem Trägheitsmoment Θ_x bezüglich der angegebenen Achse (x).
Im in der Spektroskopie üblichen Maßstab der Wellenzahlen $\tilde{\nu} = 1/\lambda$ ($W_{rot} = h\nu = hc \cdot 1/\lambda$), schreibt sich obige Formel $1/\lambda = BJ\,(J+1)$. $B = h/(8\,\pi^2\,c\Theta_x)$ (B = Rotationskonstante).
Unter Beachtung der Auswahlregeln für erlaubte optische Übergänge $\Delta J = \pm 1$, beobachtet man die Übergänge zwischen den Termen zu $1/\lambda = BJ\,(J+1) - B\,(J-1)\,J = 2\,BJ$.
Sie ergeben äquidistante Linien im Abstand $2\,B$ voneinander. Abb. 5.8.2–1 zeigt die Entstehung der meist im fernen Ultrarot gelegenen Rotationsbande. Aus dem Spektrum ist direkt die Rotationskonstante B zu entnehmen, aus der sich das Trägheitsmoment des Moleküls Θ bezüglich der jeweiligen Rotationsachse bestimmt. Kennt man die Massen der das Molekül bildenden Atome m_1 und m_2, so ist auch der Gleichgewichtsabstand r_0 der Kerne errechenbar, $\Theta = m_1 \cdot m_2 \cdot r_0^2/(m_1 + m_2)$. Die Trägheitsmomente zweiatomiger Moleküle liegen um $\approx 10^{-45}$ bis 10^{-47} kgm². Beispiele: Die Moleküle H_2, O_2 und Cl_2 haben Trägheitsmomente von 0,47; 19 bzw. $113 \cdot 10^{-47}$ kgm², ihre Gleichgewichtsabstände betragen 0,77; 1,20 bzw. $1{,}98 \cdot 10^{-10}$ m. Rotationsspektren scharfer Linien können nur in Gasen geringer Dichte beobachtet werden. Bei größeren Dichten verbreitern sich die Linien durch Stöße zwischen den Molekülen und gehen ineinander über. In Flüssigkeiten und Festkörpern fehlt die Rotationsstruktur. Viele zweiatomige Gase befinden sich bei Zimmertemperatur in angeregten Zuständen mit Rotationsquantenzahlen größer als 10.
Die Rotationsstruktur mehratomiger Moleküle ist allgemein sehr komplex. Zudem nehmen mit zunehmendem Trägheitsmoment die Linienabstände ab, so daß eine immer größere Energieauflösung erforderlich wird. Wenn man auch jede Drehbewegung in Drehungen um die drei Hauptträgheitsachsen zerlegen kann, so ist doch die Analyse der Rotation mehratomiger Moleküle äußerst kompliziert. Es kommt zu einer Kopplung der Rotation um die einzelnen Achsen, analog der Kopplungen in der Kreiselmechanik.
5.8.2.2. Schwingungsspektren. Moleküle sind immer in Schwingung, es gibt keinen schwingungsfreien Molekülzustand. Alle Schwingungsspektren haben im tiefstmöglichen Zustand noch eine *Nullpunktsenergie.* So genannt, weil sie selbst unter maximalmöglichem Energieentzug, d. h. am Nullpunkt der Temperatur, noch vorhanden ist. Sie stellt eine nicht zu beseitigende Moleküleigenschaft dar. Die Begründung dafür ist aus dem Welle-Teilchen-Dualismus resultierend, in der Heisenbergschen Unbestimmtheitsrelation zu finden (vgl. 5.5.4.4.). Nie sind Ort und Impuls eines Teilchens mit beliebiger Genauigkeit festgelegt. Ohne Schwingung wären die Atome des Moleküls gegeneinander in Ruhe und dadurch ihre Gleichgewichtsabstände beliebig genau fixiert. Dann müßte aber die Energie des Systems beliebig ungenau sein, was nicht der Fall ist. Die Energieunbestimmtheit ist ja bei stabil existierenden Molekülen zumindest wesentlich geringer als die Bindungsenergie. Die Nullpunktsenergie wurde für das H_2-Molekül zu 0,27 eV entsprechend 25,1 kJ mol^{-1} ermittelt. Die Dissoziationsenergie beträgt 4,46 eV. Um die erste Anregungsstufe des Schwingungsspektrums zu erreichen, ist etwa die doppelte Nullpunktsenergie aufzubringen. Die Energieabstufung in Schwingungsspektren ist entsprechend diesem Beispiel wesentlich größer als für Rotationsspektren. Bei Zimmertemperatur sind in der Regel Schwingungsspektren noch nicht angeregt.
In zweiatomigen Molekülen schwingen die Atome auf ihrer Verbindungsachse gegeneinander um den Gleichgewichtsabstand r_0. Die Kraft F ist wie bei dem mechanischen Modell der mit einer Feder verbundenen Kugeln proportional

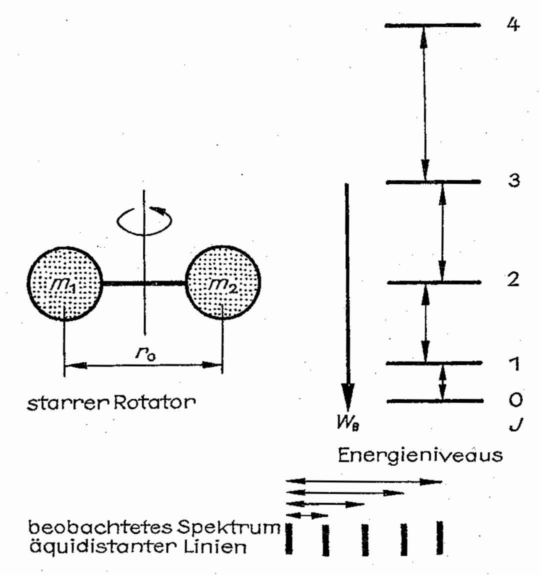


Abb. 5.8.2–1 Rotationsspektren zweiatomiger Moleküle

der Auslenkung $(r-r_0)$, $F=-k(r-r_0)$. Proportionalitätsfaktor ist die Bindungskraftkonstante k. Es ergibt sich mit diesem Kraftansatz ein parabelförmiges Potential $W_{pot}=k(r-r_0)^2/2$. Aus der Schrödinger-Gleichung errechnen sich die Energieeigenwerte der Schwingungen zu $W_v=h\nu_0(v+1/2)$ *(Harmonischer Oszillator)*, $v=0, 1, 2, \ldots$. Die Eigenfrequenz des Systems ist $\nu_0=(1/(2\pi))(k/\mu)^{1/2}$. Anstelle der reduzierten Masse μ ist auch das Trägheitsmoment $\Theta=\mu r_0^2$ einsetzbar. $v=0$ beschreibt die Nullpunktsenergie $W_0=\nu_0\cdot h/2$. Die Kraftkonstante k ist ein Maß für die Bindungsstärke, für die Festigkeit der chemischen Bindung. Damit kann aus den Schwingungsspektren die Bindungsfestigkeit beurteilt werden. So haben Doppelbindungen mehr als doppelt so große Kraftkonstanten wie Einfachbindungen. Bei Dreifachbindungen kann sich die Kraftkonstante noch einmal verdoppeln.

Der hier vorgestellte harmonische Oszillator ist für die ersten Anregungszustände eine brauchbare Näherung, berücksichtigt aber sonst den realen Potentialverlauf nicht ausreichend. Die Schwingungen der Atome sind nicht symmetrisch um den Gleichgewichtsabstand. Real steigt bei geringeren Kernabständen die Abstoßungskraft stärker an, und bei größeren Kernabständen ist die Potentialmulde ausgedehnter, und die Auslenkung geht in Höhe der Dissoziationsenergie gegen Unendlich. Abb. 5.8.2–2 verdeutlicht dies. Für das Beispiel des Wasserstoffmoleküls ist mit dem 9. Schwingungsquant schon eine Auslenkung auf das Vierfache des Gleichgewichtsabstandes von $r_0=0{,}74\cdot10^{-10}$ m erreicht, und mit dem 10. Schwingungsquant dissoziiert das Molekül. Für eine genauere Berechnung der Schwingungsmoden versucht man mathematisch den wahren Potentialverlauf möglichst gut nachzubilden. Für den Fall des Harmonischen Oszillators gilt die Auswahlregel $\Delta v=\pm1$. Die Anharmonizität verändert die Auswahlregel zu $\Delta v=\pm1$, ±2, ±3, ..., so daß unter Lichteinstrahlung mit einem Quantensprung ein hochangeregter Zustand erreicht werden kann. Die Anregung erfolgt allerdings mit abnehmender Wahrscheinlichkeit bei zunehmender Größe der Quantenzahldifferenz Δv.

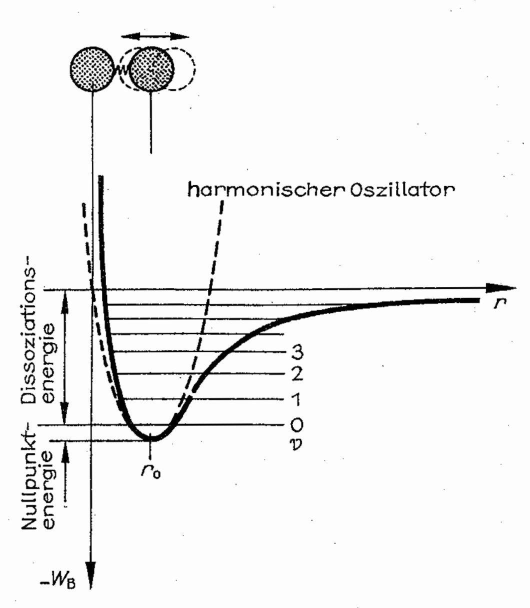
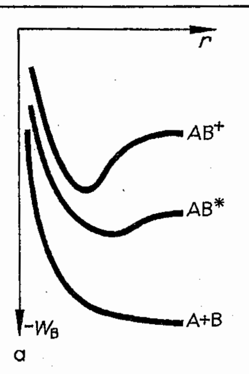


Abb. 5.8.2–2 Potentialverlauf und Lage der Schwingungsmoden

Für vielatomige Moleküle ist die Zahl der möglichen Schwingungen $3N$-6. N ist die Zahl der Atome des Moleküls. $3N$ Freiheitsgrade existieren, von denen drei für die Translation und drei für die Rotation des Moleküls in seiner Gesamtheit abzuziehen sind. Die $3N$-6-Schwingungen werden als Normalschwingungen bezeichnet. Bei den real beobachteten Schwingungen treten auch in der Regel Kopplungen zwischen den Normalschwingungen auf. Die Kopplungen können so verwickelt sein, wie bei derart vielen untereinander elastisch verbundenen Pendeln. Bei den Grundschwingungen unterscheidet man zwischen Streckschwingungen in Bindungsrichtung und Biegeschwingungen, die den Bindungswinkel ändern. Schwingungen wie Rotationen können optisch nur angeregt werden, wenn mit der Schwingung bzw. Rotation eine Änderung des elektrischen Momentes verbunden ist. Die deshalb nicht auftretenden Spektren der symmetrischen Moleküle, wie H_2, O_2 oder Cl_2, sind aber bei gleichzeitiger Anregung von Elektronen des Moleküls beobachtbar.

5.8.2.3. Elektronenanregungszustände. Wie im Atom, so kann auch im Molekül ein einzelnes Valenzelektron auf eine weiter außen gelegene Bahn gehoben werden, in einen angeregten Zustand übergehen. Die Energie zur Änderung eines Elektronenzustandes ist wesentlich größer als zur Anregung von Molekülschwingungen. Zur Änderung der Hauptquantenzahl n sind Energiebeträge notwendig, die oft über denen zur Dissoziation des Moleküls liegen. Beispielsweise erfordert der Übergang von $n=1$ auf $n=2$ im Molekül des Wasserstoffs das 2,5fache der Dissoziationsenergie. Dagegen sind Änderungen des Drehimpulses eines einzelnen Elektrons im Atom mit geringerem Energieaufwand verbunden und spielen eine wesentliche Rolle in der Molekülphysik. Die Elektronenanregungsenergien liegen oft im Gebiet der Quanten des sichtbaren Lichtes und sind deshalb von Bedeutung zur Erklärung der Farbe der Moleküle. Je nachdem, welche Farbe des Tageslichtes von einem Molekül absorbiert wird, bildet das restliche reflektierte Spektrum die Farbe des Moleküls.

Zur Beschreibung der Elektronenzustände im Molekül verwendet man die zum Atom analogen Bezeichnungen. Bei zweiatomigen Molekülen präzessiert der Bahndrehimpuls l der Elektronen unter dem Einfluß des axialen Kernfeldes um die Verbindungslinie beider Kerne. Die Komponenten in Richtung dieser Bindungsachse sind in der Bedeutung der l_z-Komponente bzw. der Magnetquantenzahl m analog, sie werden mit λ bezeichnet. λ hat auch die Mannigfaltigkeit $2l+1$, die möglichen Werte $\lambda=\pm1$, $\pm(l-1)$, $\pm(l-2), \ldots, 0$. Für die Elektronenzustände $\lambda=0, 1, 2, 3, \ldots$ sind die griechischen Buchstaben $\sigma, \pi, \delta, \varphi, \ldots$ (Sigma, Pi, Delta, Phi) eingeführt. Daraus resultiert als Beispiel die Molekülelektronenbezeichnung 3 dπ, für ein Elektron mit $n=3$, $l=2$ und der Drehimpulskomponente in Bindungsrichtung $\lambda=1$. Bei mehreren Molekülelektronen wird die resultierende Gesamtbahndrehimpulsquantenzahl auch in Analogie zum Atom mit großen Buchstaben gekennzeichnet. Wegen des starken axialen Feldes addieren sich die λ-Werte: $\Lambda=\Sigma_i\lambda_i$, für $\Lambda=0, 1, 2, 3$ schreibt man $\Sigma, \Pi, \Delta, \Phi$. Der Grundzustand fast aller Moleküle ist zu $\Lambda=0$ abgesättigt, wird also mit Σ bezeichnet sein. Der Gesamtspin S bildet sich aus der Summe der einzelnen Elektronenspins und präzessiert um die Bindungsachse. Er bewirkt die Multiplizität $2S+1$ der Λ-Terme. Somit läßt sich das H_2-Molekül



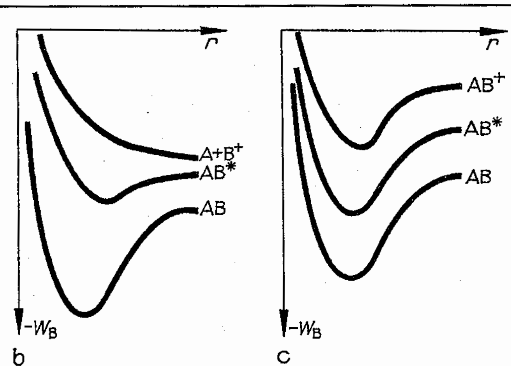


Abb. 5.8.2–3 Potentialkurven verschiedener Moleküle im Grundzustand (AB), dissoziiert (A + B), angeregt (AB^*) und ionisiert (AB^+)

durch die Bezeichnung $(1\,s\,\sigma)^2\ {}^1\Sigma_0$ darstellen. Sie besagt, daß zwei Elektronen (Index 2 zur Klammer) im Zustand $n = 1$, $l = 0$ und $\lambda = 0$ das Molekül mit dem Gesamtdrehimpuls $\Lambda = 0$, dem Gesamtspin $S = 0$ und der Multiplizität 1 (nicht entartet) bilden. Bei Elektronenanregung wird eher ein Elektron über viele Stufen angeregt, wobei bei den Rydberg-Serien ähnliche Linienfolgen entstehen, als daß mehrere Elektronen gleichzeitig angeregt werden. Wie weit die Molekülanregung gehen kann, entscheidet auch die Bindung. Abb. 5.8.2–3 zeigt mögliche Potentialformen bei Elektronenanregung im Molekül. Je nachdem, ob Rydberg-Serien bis in das Seriengrenzkontinuum (Ionisation) reichen, wie beim Molekül des Wasserstoffs (Abb. 5.8.2–3 c), nur das Molekülion stabil ist wie im Falle des Heliums (Abb. 5.8.2–3 a), oder ob das Molekülion instabil ist (Abb. 5.8.2–3 b), so bilden sich die Potentialkurven aus.

In der Regel findet man keine reinen Elektronenanregungsspektren. Gewöhnlich werden Schwingungszustände des Moleküls mit angeregt. Auch die Rotationsniveaus sind dann nicht im Grundzustand, so daß die Elektronenterme von Schwingungsniveaus überlagert sind und auf diesen noch die Linien der Rotationszustände liegen. Die Linien aller Rotationszustände eines Schwingungsniveaus werden als Bande bezeichnet, ein Bandensystem (auch Bandengruppe) besteht aus den Banden aller Schwingungsniveaus auf einem Elektronenterm.

5.8.3. Kombiniertes Auftreten verschiedener Molekülanregungsarten

5.8.3.1. Rotationsschwingungsspektren. Die den Schwingungszuständen überlagerten Rotationsniveaus ändern sich auch bei Änderung des Schwingungszustandes. Für harmonische Schwingungen gilt die Auswahlregel $\Delta v = \pm 1$, für unharmonische Schwingungen sind dies die Übergänge mit der größten Intensität. Im Beispiel der Abb. 5.8.3–1 ist $\Delta v = +1$. Die möglichen Änderungen der Rotationsniveaus sind $\Delta J = \pm 1$. Da immer auch viele Rotationsniveaus angeregt sind, beobachtet man die im Bild eingezeichneten beiden Zweige einer Rotationsschwingungsbande. Der linke Zweig gehört zu $\Delta J = +1$ und der rechts zu $\Delta J = -1$. Die Übergänge $\Delta J = 0$ sind verboten, so daß auch die Linie des reinen Schwingungsüberganges in der Bande nicht beobachtet wird. Die Linienabstände in den Zweigen sind etwa konstant und gleich $2B$, wie bei reinen Rotationsspektren. Für genauere Betrachtungen ist zu berücksichtigen, daß das in B enthaltene Trägheitsmoment nicht vom Schwingungszustand unabhängig ist. Es vergrößert sich etwas mit Vergrößerung der Schwingungsquantenzahl v. Auch sind Schwingungen und Rotationen in Wechselwirkung, und für genaue Betrachtungen ist die einfache Addition beider Terme durch einen Wechselwirkungsanteil zu korrigieren.

5.8.3.2. Bandenspektrum bei Elektronenanregung. Die Banden der den Schwingungstermen überlagerten Rotationslinien unterscheiden sich bei gleichzeitigem Elektronensprung erheblich von dem eben diskutierten einfacheren Fall der Übergänge ohne Elektronenanregung. Neben dem positiven und negativen Zweig der Rotationsschwingungsbande ist auch der Null-Zweig möglich, da nun die Auswahlregel $\Delta J = 0$, ± 1 gilt. Desweiteren ändern sich die Potentialkurven durch einen Elektronensprung erheblich, wodurch auch die nun zwischen diesen Kurven erfolgenden Übergänge stark verändert sein können. Durch die Änderung des Gleichgewichtsabstandes, der Bindungskraftkonstanten und der Trägheitsmomente sind die Rotationsbande nicht mehr einfach zu analysieren. Die Abstände der

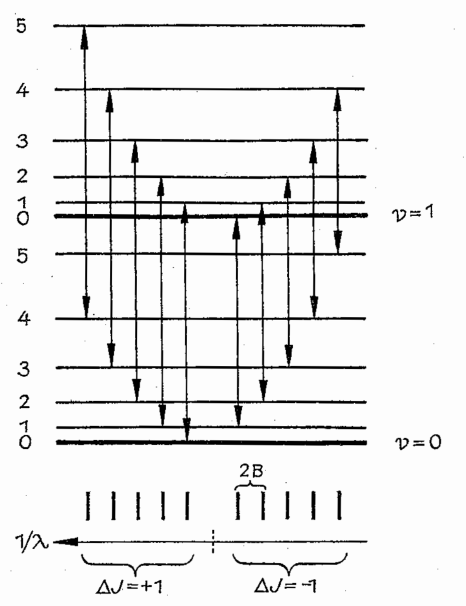


Abb. 5.8.3–1 Termschema und Linienfolge einer Rotationsschwingungsbande (v = Schwingungs-, J = Rotationsquantenzahl)

Linien können sich innerhalb einer Bande sehr verändern und die Frequenzen der Linien mit wachsender Rotationsquantenzahl sogar erst steigen und dann fallen, so daß die Folge der Linien ihre Richtung umkehrt und eine kompliziert zu entschlüsselnde Linienhäufung zustande kommt. Eine übersichtliche Darstellung der Termfolge einer Bande ermöglicht das Fortrant-Diagramm. In der zweidimensionalen Anordnung wird das Frequenzspektrum nach der Rotationsquantenzahl J zerlegt. Jeder Linie des Spektrums entspricht ein Punkt in der J- über $1/\lambda$-Auftragung. Abb. 5.8.3–2 zeigt eine Elektronenbande und das zugehörige Fortrant-Diagramm. Die Zweige der Bande können auch in der anderen Richtung gekrümmt auftreten, je nach der relativen Größe der Rotationskonstanten B und der Gleichgewichtsabstände r_0 im Molekül vor und nach dem Elektronensprung. Durch Analyse der Spektren kann aus den Umkehrpunkten der Zweige über die relative Lage dieser Größen eine Aussage gemacht werden. Kehrt der Zweig zu größeren Frequenzen hin um (Umkehr im $\Delta J = +1$-Zweig), so ist der Gleichgewichtsabstand im oberen Zustand der beiden Schwingungsterme größer als im unteren und die Rotationskonstante im unteren Zustand größer als im oberen. Umgekehrt, kehrt ein Zweig zu geringeren Frequenzen hin um (was dann mindestens im $\Delta J = -1$-Zweig erfolgt), so vertauschen sich auch die Relationen zwischen den Gleichgewichtsabständen und Rotationskonstanten. Das wird verständlich, wenn man die Wellenzahlen des Bandenspektrums genauer betrachtet: $1/\lambda = 1/\lambda_0 + B' J'(J'+1) - B'' J''(J''+1)$ *(Elektronenbande).* $1/\lambda_0$ kennzeichnet die Lage des Spektrums durch den Elektronen- und Schwingungsquantensprung, auf die sich das Rotationsspektrum addiert. Wie üblich sind die energetisch höher angeregten Zustände mit einem Strich und die zu dem niederen Schwingungsquant gehörenden Zustände mit zwei Strichen gekennzeichnet. Gemäß der Auswahlregel für den Rotationsquantensprung $\Delta J = 0, \pm 1$ ergeben sich die drei Zweige der Bande:

$\Delta J = 0$-Zweig:

$$1/\lambda = 1/\lambda_0 + (B' - B'') J(J+1); \; J = 1, 2, 3, 4, \ldots$$
$$= 1/\lambda_0 + (B' - B'') J^2 + (B' - B'') J$$

$\Delta J = -1$-Zweig:

$$1/\lambda = 1/\lambda_0 + B' J(J-1) - B'' J(J+1)$$
$$= 1/\lambda_0 + (B' - B'') J^2 - (B' + B'') J;$$
$$J = 0, 1, 2, 3, \ldots$$

$\Delta J = +1$-Zweig:

$$1/\lambda = 1/\lambda_0 + B'(J+1)(J+2) - B'' J(J+1)$$
$$= 1/\lambda_0 + (B' - B'') J^2 + (3B' - B'') J + 2B';$$
$$J = 0, 1, 2, 3, \ldots$$

Das Franck-Condon-Prinzip formuliert Regeln, welche der

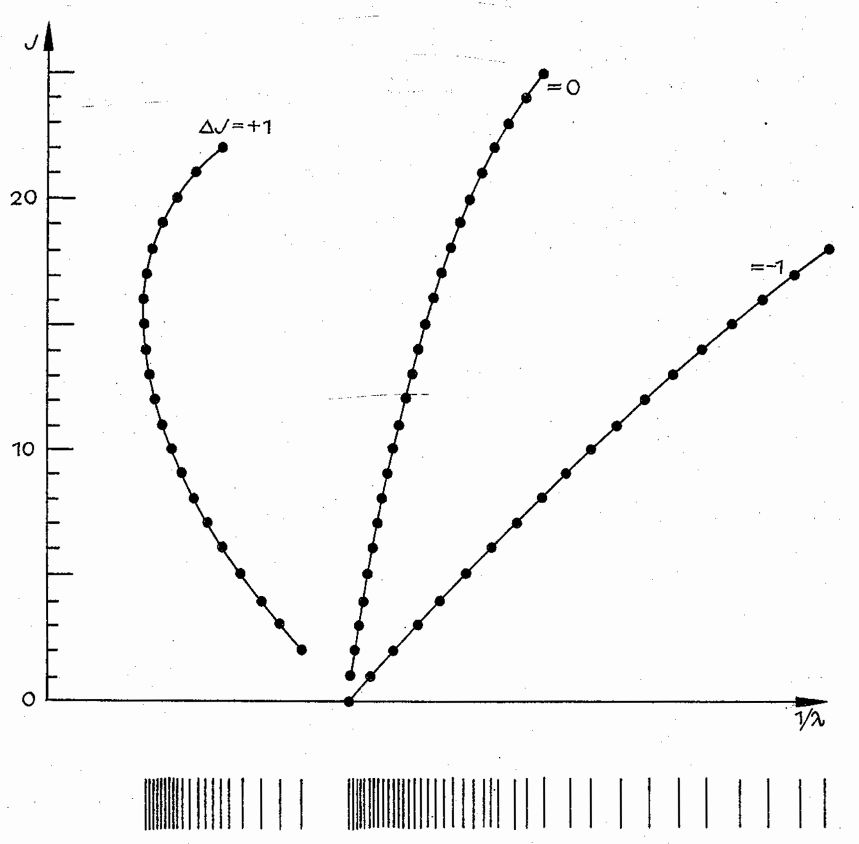


Abb. 5.8.3–2 Spektrum und Fortrant-Diagramm der drei Zweige einer Bande im Rotationsschwingungsspektrum bei gleichzeitiger Anregung eines Molekülelektrons

vielen möglichen Kombinationen von gleichzeitigen Elektronen- und Schwingungsquantensprüngen intensitätsstark erfolgen, d.h. bevorzugt stattfinden. Grundlage des Prinzips ist der Sachverhalt, daß Elektronensprünge von einem Orbital zum anderen sehr schnell erfolgen (10^{-15} s), sehr schnell gegen die Bewegungen der Atome des Moleküls. Insbesondere werden sich während eines Elektronensprunges der Gleichgewichtsabstand der Atome zueinander und ihre Geschwindigkeiten nicht wesentlich verändern. Das bevorzugt in etwa senkrechte Übergänge in den üblichen Darstellungen des Potentialverlaufes, wie sie Abb. 5.8.3–3 zeigt und von diesen wiederum nur die in den Schwingungsumkehrpunkten. Eine Ausnahme bildet der Zustand $v = 0$. Hier beginnen und enden bevorzugte Übergänge im Mittelteil. Die Regeln begründen sich quantenmechanisch durch die in den Umkehrpunkten (und im v_0-Zustand im mittleren Teil) besonders großen Aufenthalts- und damit auch Übergangswahrscheinlichkeiten.
Aus der Analyse der mit größter Intensität auftretenden Übergänge kann man Schlüsse über die relative Lage der Potentialkurven der beiden Elektronenzustände ziehen. Insbesondere ergibt sich, ob die Elektronenanregung die Molekülbindung verfestigt (r_0 wird kleiner), oder, wie allgemein zu erwarten, die Molekülbindung gelockert wird (r_0 wird größer).

5.8.3.3. Dissoziation und Bildung von Molekülen unter Lichtabsorption bzw. -emission. Die zum Zerfall eines Moleküls in seine Atome notwendig aufzubringende Energie (Dissoziationsenergie), die gleich der bei der Molekülbildung freiwerdenden Energie ist, ist eine wesentliche Größe zur Beschreibung des Moleküls, sie bestimmt seine Stabilität. Es ist nicht möglich, durch Lichteinstrahlung das Molekül in einen so hohen Schwingungs- und Rotationsquantenzustand zu versetzen, daß es in seine Atome zerfällt. Es läßt sich nicht einfach durch sukzessive Erhöhung der Frequenz des eingestrahlten Lichtes ermitteln, daß z.B. das Molekül zwischen dem 9. und 10. Schwingungsquant, etwa bei der Rotationsquantenzahl $v = 27$ dissoziiert. Das Franck-Condon-Prinzip verbietet es. Es wären Übergänge entsprechend den durchgekreuzten Pfeilen in Abb. 5.8.3–4. Ein Schwingungsquant nach dem anderen zu übertragen und den Anregungszustand Schritt für Schritt zu erhöhen, gelingt auch nicht. Ein Anregungszustand ist nicht langlebig genug, das Eintreffen eines weiteren Lichtquantes zu erleben.
Dennoch ist der Molekülzerfall unter Lichteinwirkung *(Fotodissoziation)* möglich. Er kann über einen angeregten Elektronenzustand erfolgen. Für das angeregte Molekül gilt eine andere Potentialkurve, die so günstig liegen kann, daß der Übergang in das Dissoziationskontinuum dort gemäß dem Franck-Condon-Prinzip wahrscheinlich ist. (Diese Möglichkeit ist in Abb. 5.8.3–4 dargestellt.) Es ist ein Umweg über Energiezustände, die die gesuchte Dissoziationsenergie W_{Diss} des Moleküls im Grundzustand weit übersteigt. Daß bei diesem Zerfall ein Atom in einem angeregten Zustand zurückbleibt, ist verständlich. Die auf das Molekül übertragene Energie W_1 übersteigt um W_2 die notwendige Dissoziationsenergie W_{Diss} und muß, wenn nichts anderes beobachtet wird, bei den getrennten Molekülbausteinen verblieben sein. Die Energie W_2 ist aus Messungen zu ermitteln, so daß auf diesem Wege W_{Diss} ermittelt werden kann.
Die Fotodissoziation ist nicht die einzige Möglichkeit, Moleküle zu zerlegen. Das geschieht einfacher durch Erwärmen eines Gases der Moleküle. Sie übertragen einander Energie durch Stöße, für die natürlich die optischen Auswahlregeln und das Franck-Condon-Prinzip nicht gelten. W_{Diss} ist aber dabei nicht ohne weiteres etwa aus der Gastemperatur zu ermitteln, auch nicht als Mittelwert. Es fehlen Aussagen über die Anregungszustände, in denen sich das Molekül im Moment des Zerfalles befindet.

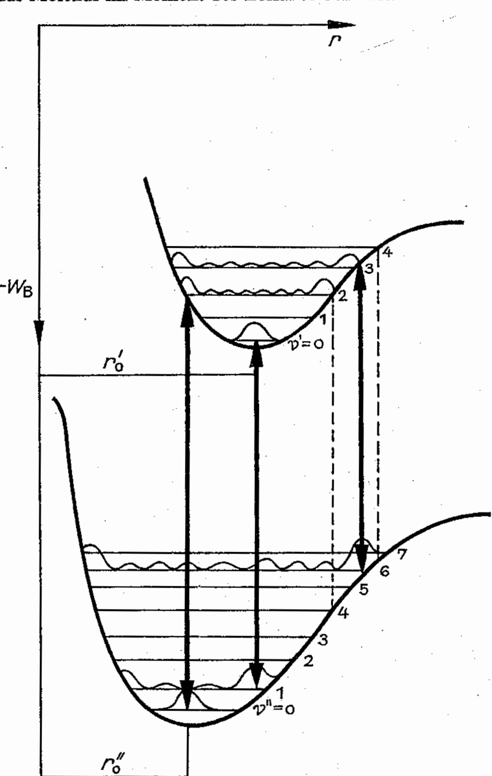


Abb. 5.8.3–3 Franck-Condon-Prinzip zu den wahrscheinlichsten Übergängen bei gleichzeitiger Änderung von Schwingungs- und Elektronenanregungszustand. Mit den Wellenlinien sind die Aufenthaltswahrscheinlichkeiten in den Zuständen zu bevorzugten Übergängen eingezeichnet

Die Molekülbildung in Umkehrung des bisher Diskutierten, d. h. unter Aussendung von Licht, ist ganz analog zu verstehen. Wenn zwei Atome mit nur geringer kinetischer Energie sich einander nähern, begegnen sie sich im Potentialbild der Abb. 5.8.3–4 in Höhe des auslaufenden Zweiges der unteren Potentialkurve. In einen tiefer gelegenen, bindenden Zustand gelangen die Atome nur durch Energieabgabe. Die Aussendung eines Lichtquantes würde die Umkehrung eines Überganges entlang der durchkreuzten Pfeile bedeuten und mit gleicher Begründung verboten sein. Die Molekülbildung über einen angeregten Zustand gemäß der oberen Potentialkurve ist möglich, muß aber sinngemäß unter Emission eines Strahlungsquantes der Energie W_1 erfolgen. Dieser „Strahlungsrekombination“ genannte Vorgang verlangt den Stoß eines Atoms im Grundzustand mit einem Atom in einem angeregten Zustand. Die Strahlungsrekombination ist in den entsprechenden erhitzten Gasen beobachtbar.

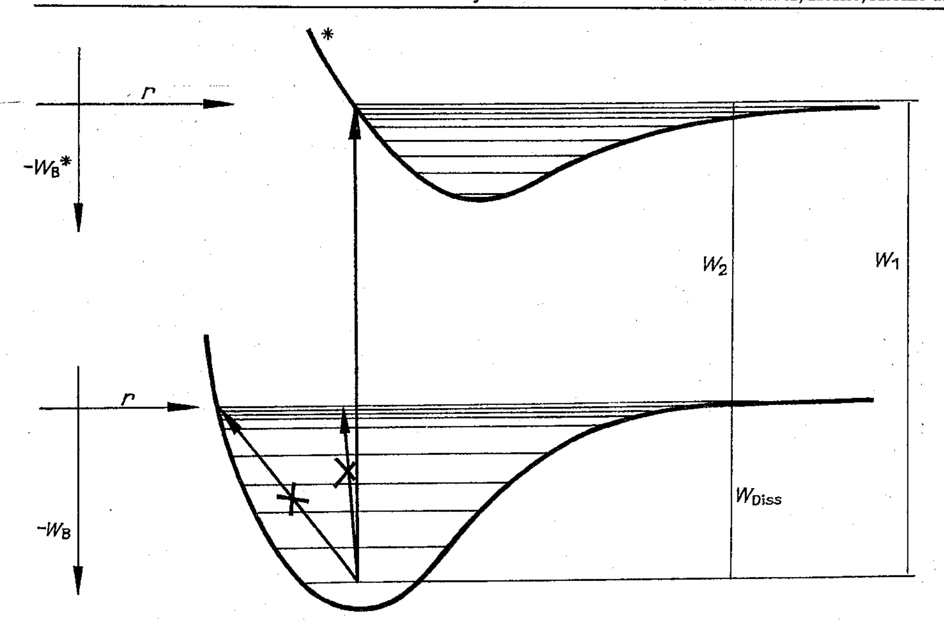


Abb. 5.8.3-4 Zum Verbot der direkten Fotodissoziation (durchkreuzte Pfeile) und die Möglichkeit, über Elektronenanregung (Potentialkurve mit Stern) zu dissoziieren

5.8.3.4. Raman-Effekt. Bestrahlt man Moleküle mit monochromatischem Licht, so wird in vielen Fällen nicht nur Streulicht der eingestrahlten Frequenz beobachtet, sondern es treten noch rechts und links davon weitere, wenn auch intensitätsschwache Linien auf. Dieser Effekt wird nach einem der Entdecker des schon von SMEKAL vorhergesagten Phänomens *Raman-Effekt* genannt.

Im Feld einer eingestrahlten Welle der Kreisfrequenz ω_0 wird ein mit gleicher Frequenz sich änderndes Dipolmoment M_i induziert, $M_i = \alpha E_0 \cos \omega_0 t$ (α = Polarisierbarkeit). Derart realisiert sich die normale Lichtstreuung (Rayleigh-Streuung). Ist, genauer betrachtet, die Polarisierbarkeit nicht konstant, sondern von der Auslenkung der Atome im Molekül abhängig, so ist folgender Ansatz gerechtfertigt: $\alpha = \alpha_0 + \alpha_1 \cos \omega_s t$. Für ω_s sind die Kreisfrequenzen der angeregten Rotations- oder Schwingungszustände, in denen sich die Moleküle zur Zeit der Lichteinstrahlung gerade befinden, einzusetzen. Daraus ergibt sich das induzierte Dipolmoment zu

$$M_i = (\alpha_0 + \alpha_1 \cos \omega_s t)\, E_0 \cos \omega_0 t$$
$$= \alpha_0 E_0 \cos \omega_0 t$$
$$+ 1/2\, \alpha_1 E_0 [\cos(\omega_0 - \omega_s)\, t + \cos(\omega_0 + \omega_s)\, t]\,.$$

Das erste Glied erzeugt, wie schon erwähnt, das Streulicht gleicher Frequenz. Die um die Frequenzen der Rotations- und Schwingungszustände zu größeren und kleineren Werten hin verschobenen Linien werden Raman-Linien genannt. Die Linienfolge zu geringeren Frequenzen hin trägt die Bezeichnung Stookessche Linien und zu höheren Frequenzen hin die Bezeichnung Antistookessche Linien. Im ersten Fall überträgt das eingestrahlte Licht noch Energie zur Anregung von Rotationszuständen, im letzten übernimmt das Streulicht die Energie von Rotationszuständen. Nicht alle Moleküle sind „Raman-aktiv“ und besitzen diese Satellitenlinien.

Der Raman-Effekt ist wesentlich zur Untersuchung der Rotations- und Schwingungszustände von Molekülen ohne permanentes Dipolmoment, bei denen eine optische Anregung nicht möglich ist. Bei der Analyse der Anregungszustände über den Raman-Effekt befinden sich die Linien dicht neben der Linie des eingestrahlten Lichtes, sind dorthin „transformiert“. Dadurch kann man sich den zur Analyse der Rotationsstruktur eines Moleküls zu untersuchenden Frequenzbereich, eventuell nach dem vorhandenen Spektrometer höchster Auflösung, aussuchen. Heutzutage ist die Raman-Spektroskopie unter Verwendung des extrem monochromatischen und intensitätsstarken Laserlichtes mit großem praktischen Nutzen entwickelt worden. Mit ihr können beispielsweise aus einer Erdumlaufbahn Luft und Wasser auf bestimmte Moleküle hin analysiert werden.

6. Kondensierte Materie

Festkörper, feste Stoffe, haben als Werkstoffe und Gebrauchsgegenstände für die Menschen seit der Menschwerdung große Bedeutung. Die Erschließung natürlicher (Holz, Knochen, Stein) und künstlicher Werkstoffe (Bronze, Eisen, Glas, Keramik) hat sich wesentlich auf die Lebensweise und Produktivität der Menschen ausgewirkt. Dies gilt auch für die heutige Zeit, in der z. B. neue Materialien die Entwicklung der Mikroelektronik und damit eine Revolutionierung der Technologie in vielen Industriezweigen ermöglichten.

Die Suche nach neuen Werkstoffen wird durch die Anforderungen an die Eigenschaften der Werkstoffe bestimmt. Solche Eigenschaften waren zunächst Härte, Verschleißfestigkeit, Biegsamkeit, Wärmedämmung, Hitzebeständigkeit, Wärmeausdehnung, d. h. mechanische und thermische Eigenschaften; später kamen optische, magnetische und elektrische Eigenschaften hinzu.

In der Festkörperphysik werden diese Eigenschaften systematisiert, Beziehungen zwischen verschiedenen Eigenschaften und Abhängigkeiten der Eigenschaften von Zustandsgrößen, wie Temperatur, Druck und chemischer Zusammensetzung, aufgedeckt. Als der atomare Aufbau der Stoffe und die regelmäßige Anordnung der Atome im Kristallgitter erkannt worden waren, wurden in der Festkörperphysik bedeutende Anstrengungen unternommen, die Eigenschaften des Festkörpers aus seiner atomaren Struktur heraus zu verstehen. Neben die makroskopische, die Kontinuumsbeschreibung, trat die mikroskopische, die atomare Beschreibung. Die mikroskopische Beschreibung erforderte die Entwicklung der theoretischen Grundlagen der Atomphysik, der Quantentheorie, und die Entwicklung einer umfangreichen Analysenmeßtechnik, in der heutzutage vielfach Großgeräte eingesetzt werden.

Während die makroskopische Beschreibung dominiert, wo große Mengen von Material eingesetzt werden, wo polykristallines Material verwendet wird, ist die mikroskopische Beschreibung dort unerläßlich, wo mit Einkristallen oder mit kleinen Bauelementen gearbeitet wird. Da dies in der Mikroelektronik der Fall ist und diese Technik wegen ihres geringen Materialeinsatzes, wegen des geringen Energieverbrauchs, wegen kurzer Schaltzeiten usw. immer stärker zum Einsatz kommt, ist in der Festkörperphysik eine intensive Kopplung zwischen Grundlagenforschung und Technologie entstanden.

Die Eigenschaften des Festkörpers, insbesondere seine elektrischen, magnetischen und optischen Eigenschaften, hängen wesentlich von der geometrischen Struktur, der Anordnung der Atome, ab. Deswegen zielen die Züchtungen darauf hin, einerseits perfekt gebaute Einkristalle herzustellen und andererseits Defekte in definierter Weise einzubauen. Auch die stöchiometrische Zusammensetzung soll mit hoher Genauigkeit eingehalten werden, und Fremdstoffe dürfen nur in verschwindenden Konzentrationen auftreten. Hohe Anforderungen in dieser Hinsicht werden insbesondere an die Materialien der Halbleitertechnik und Mikroelektronik gestellt.

Die sich einstellende Kristallstruktur und die Eigenschaften des Kristalls werden durch die Wechselwirkung der Valenzelektronen der Atome und durch die Art der Zustände, die diese Elektronen im Kristall einnehmen, bestimmt. Daher spielt bei der mikroskopischen Betrachtung die experimentelle und theoretische Untersuchung der Elektronenstruktur der Festkörper eine dominierende Rolle. Die in den letzten Jahrzehnten erzielten Fortschritte in der Kenntnis der Elektronenstruktur der Festkörper haben die stürmische Entwicklung in der Halbleiterphysik und in der Mikroelektronik erst ermöglicht.

Der Inhalt des Teiles Festkörperphysik lehnt sich an die genannten Gesichtspunkte an. Der Diskussion der Struktur, der elektronischen Eigenschaften und der Gitterschwingungen des perfekt gebauten Kristalls schließt sich die Diskussion des Einflusses von Defekten (Leerstellen, Fremdatome, Versetzungen) und von Oberflächen an. Bei der Diskussion der mechanischen, elektrischen, magnetischen und optischen Eigenschaften steht die makroskopische Beschreibungsweise im Vordergrund. Bezüglich der thermischen Eigenschaften kann außer bei der Kristallzüchtung im wesentlichen auf die Wärmelehre (2.6.) zurückgegriffen werden. Die Diskussion der Wechselwirkung des Festkörpers mit Teilchen und Feldern gibt einen Einblick in die Vielfalt der Analysenmethoden.

Die Sprache der Physik ist die Mathematik. Das kommt auch in der Festkörperphysik zum Ausdruck. Neben der mathematischen Formulierung steht aber auch die verbale Erläuterung und die Veranschaulichung von Begriffen und Beziehungen. Dies hilft, sich zu orientieren, Vorstellung über den Aufbau des Festkörpers und über den Zusammenhang zwischen der Struktur und den Eigenschaften des Festkörpers zu entwickeln.

6.1. Struktur von Festkörpern

Im Unterschied zu Flüssigkeiten und Gasen setzen Festkörper einer Scherdeformation beträchtlichen Widerstand entgegen. Als fest bezeichnet man eine Substanz, deren Zähigkeit $\eta = \sigma/(2\dot{\varepsilon})$ groß gegen $10^{14}\,\mathrm{Pa \cdot s}$ ist, wobei σ eine Zugspannung und $\dot{\varepsilon}$ die resultierende Dehnungsgeschwindigkeit bedeuten. Die Bestandteile des Festkörpers befinden sich offenbar im Zustand stabilen (oder metastabilen) mechanischen Gleichgewichtes. Das geht einher mit einer charakteristischen räumlichen Anordnung der Bestandteile, der *Struktur*.

6.1.1. Bestandteile und Gefüge der Festkörper

Festkörper lassen sich durch Kondensation aus gasförmigen Elementen entstanden denken. Ihre Bestandteile sind demnach die gleichen wie die der freien Atome, also *Atomkerne* und *Elektronen*. Für die Mehrzahl der Festkörpereigenschaften ist der innere Aufbau der Atomkerne unerheblich. Interferenzen mit Röntgenstrahlen und später auch solche mit Elektronen, Neutronen, Protonen u. a. erbrachten den

Beweis, daß die Mehrzahl der Festkörper im thermodynamischen Gleichgewicht Kristalle sind. Ihr wesentliches Merkmal ist eine Anordnung der Atomkerne und Elektronen, die *Kristallstruktur*, die in jeder der drei Dimensionen periodisch verläuft (Abb. 6.1.1–1). Die strenge Periodizität ist allerdings ein niemals realisierter Grenzfall (Idealkristall). Abweichungen davon werden als *Kristallbaufehler* bezeichnet. Oft ist deren Konzentration gering. Wo das nicht zutrifft, kann eine andere Definition des kristallinen Zustandes zweckmäßiger sein, etwa die, nach der nur die Verknüpfung unmittelbar benachbarter Atome oder Atomgruppen festgelegt wird.

Fehlt die Periodizität der Atomanordnung in allen drei Dimensionen, liegt der Grenzfall des *nichtkristallinen* oder *amorphen Festkörpers* vor, z. B. silikatische und metallische Gläser, Teer, Harze. Zur Kennzeichnung der mittleren Umgebung eines Teilchens wird die *Paarkorrelationsfunktion* $g(r)$ (vgl. 4.9.5.) angewendet, in der $4\pi r^2 \mathrm{d}r \varrho_0 g(r)$ die Zahl der sich im Abstand r von ihm im Volumenelement $4\pi r^2 \mathrm{d}r$ befindlichen Teilchen und ϱ_0 die mittlere Teilchendichte des Systems bedeuten (Abb. 6.1.1–2). Als Beschreibungsgrundlage der räumlichen Anordnung kann eine idealisierte amorphe Bezugsstruktur gewählt werden, die nach BERNAL in der *ungeordneten dichten Packung* (UDP) starrer Kugeln besteht (Abb. 6.1.1–3).

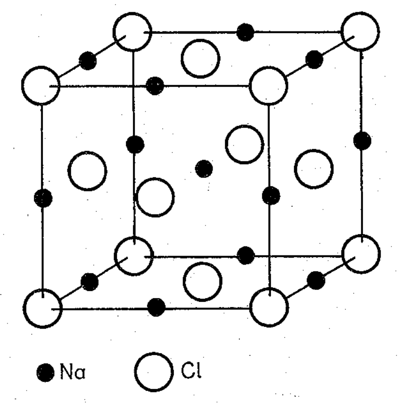


Abb. 6.1.1–1 Ausschnitt aus der Anordnung von Natrium- und Chlorionen im Kochsalzkristall. Die Kreise sollen deren Positionen symbolisch anzeigen

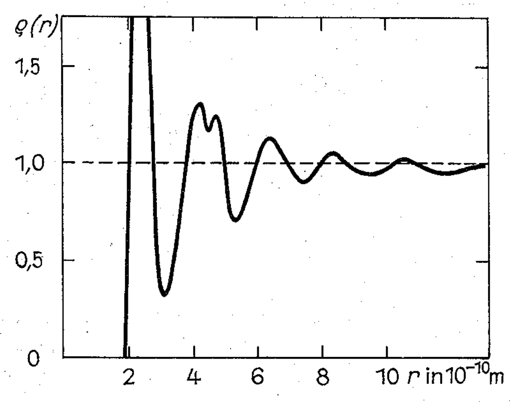


Abb. 6.1.1–2 Radialverteilungsfunktion $\varrho(r)$ von amorphem $Ni_{76}P_{24}$ (nach CARGILL). Das erste Maximum von $\varrho(r)$ liegt bei etwa $r \approx 1$ Atomdurchmesser

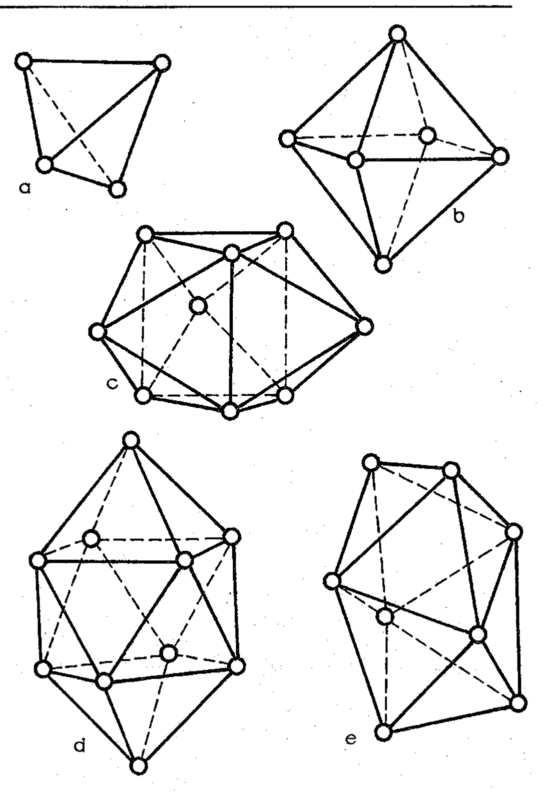


Abb. 6.1.1–3 Die fünf Bernalschen Polyeder zur Beschreibung der UDP starrer Kugeln: (a) Tetraeder, (b) Oktaeder, (c) durch drei Halboktaeder abgeschlossenes trigonales Prisma, (d) durch zwei Halboktaeder abgeschlossenes Archimedisches Antiprisma, (e) tetragonales Dodekaeder (nach POLK)

Erstreckt sich eine bestimmte kristalline Struktur durch den gesamten Festkörper, wird dieser als *Einkristall* bezeichnet. Ist das nicht der Fall, liegen also unterschiedlich orientierte Bereiche der gleichen und/oder verschiedener Strukturen vor, spricht man von einem *Vielkristall*. Die Gesamtheit dieser Bereiche, einschließlich möglicher nichtkristalliner, bildet das Gefüge des Festkörpers.

6.1.2. Geometrie des Kristallgitters

Wird in einer kristallinen Struktur, von Störungen abgesehen, das per definitionem auftretende periodisch wiederholte Motiv (kleinste Gruppe von Ionen, Atomen oder Molekülen, als deren räumliches Nebeneinander der Festkörper anzusehen ist) derart durch einen Punkt ersetzt, daß eine Anordnung mit für jeden Punkt identischer Umgebung entsteht, erhält man das Gitter der Kristallstruktur oder kurz Kristallgitter. Enthält das Motiv mehr als einen Atomkern, heißt es *Basis des Gitters*.

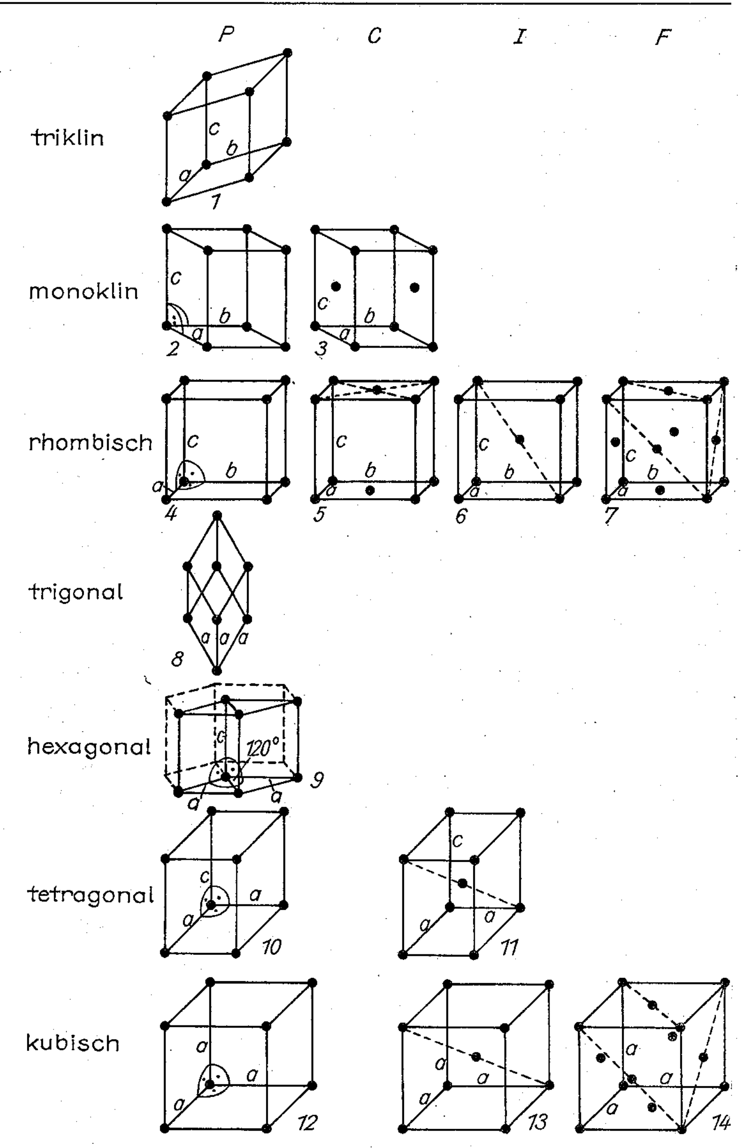


Abb. 6.1.2–1 Konventionelle Elementarzellen der 14 Bravais-Gitter. Die Symbole *P, C, I* und *F* bedeuten primitive, einseitig-flächenzentrierte, innenzentrierte und allseitig-flächenzentrierte Elementarzelle. Links sind die zugehörigen Kristallsysteme angegeben

6.1.2.1. Elementarzelle. Die räumliche Lage der Gitterpunkte kann durch Vektoren $\boldsymbol{R} = u\boldsymbol{a}_1 + v\boldsymbol{a}_2 + w\boldsymbol{a}_3$ gekennzeichnet werden. Die Vektoren $\boldsymbol{a}_i$ sind zwar beliebig wählbar, im Interesse einer einfachen Verständigung werden aber meist bestimmte Vereinbarungen befolgt (Abb. 6.1.2–1). Das von den $\boldsymbol{a}_i$ aufgespannte Parallelpipéd heißt *Elementarzelle des Gitters*, da aus ihm das gesamte Gitter durch periodische Wiederholung dargestellt werden kann. Alle denkbaren Kristallstrukturen lassen sich nach A. Bravais auf eines von insgesamt 14 verschiedenen Gittern zurückführen. In Abb. 6.1.2–1 sind die Beträge der $\boldsymbol{a}_i$, die Gitterkonstanten, durch gleiche Buchstaben bezeichnet, wenn sie symmetrisch übereinstimmen. Weisen die Winkel zwischen den $\boldsymbol{a}_i$ feste Werte auf, ist dies ebenfalls angegeben.

6.1.2.2. Rationale Kristalleigenschaften. Punkte, Geraden und Ebenen im Kristall heißen rational, wenn ihre Lage durch Gitterpunkte festgelegt ist. Rationale Punkte und Geraden können durch einen Vektor $\boldsymbol{R} = u\boldsymbol{a}_1 + v\boldsymbol{a}_2 + w\boldsymbol{a}_3$ mit rationalen Zahlen (den Indizes) u, v, w beschrieben werden. Es ist üblich, für *Punkte* kurz *uvw* und für *Geraden* [*uvw*] bzw. ⟨*uvw*⟩ zu schreiben. Spitze Klammern stehen für die Gesamtheit symmetrisch gleichwertiger Gittergeraden. Rationale Ebenen, die *Netzebenen*, beschreibt man durch $\boldsymbol{H} \cdot \boldsymbol{r} = hx_1 + kx_2 + lx_3 = 1$, wobei sich die Komponenten des Ortsvektors $\boldsymbol{r} = \sum_i x_i \boldsymbol{a}_i$ auf das Koordinatensystem der $\boldsymbol{a}_i$ beziehen und die sogenannten *Millerschen Indizes h, k, l* der Netzebene rationale Zahlen sind. $\boldsymbol{H} = (h\boldsymbol{a}_1^* + k\boldsymbol{a}_2^* + l\boldsymbol{a}_3^*)$ steht senkrecht auf der Netzebene *hkl* und kann durch die

Basisvektoren des *reziproken Gitters* (vgl. 6.3.1.) einfach dargestellt werden. Eine bestimmte Netzebene wird kurz durch (*hkl*), die Gesamtheit gleichwertiger durch {*hkl*} bezeichnet. Geometrisch bedeuten die Millerschen Indizes die reziproken Werte der Längen (in Einheiten a_i), die die Ebene auf den zugehörigen Koordinatenachsen abschneiden. Parallele Geraden und Ebenen gelten in jedem Falle als gleichwertig.

6.1.3. Symmetrieeigenschaften des Kristalles

6.1.3.1. Symmetrieoperationen. Neben der *Translationssymmetrie* (vgl. 6.1.1.) besitzt die Anordnung der Bausteine des Kristalles i. allg. noch weitere Symmetrieeigenschaften, die sich in den physikalischen und chemischen Eigenschaften widerspiegeln. Bei ungehindert wachsenden Kristallen zeigt z. B. die äußere Form teilweise diese Symmetrieeigenschaften. Die Zahl der überhaupt möglichen Symmetrieoperationen, durch die der Kristall in sich selbst überführt werden kann, ist begrenzt.

6.1.3.2. Symmetrieelemente. Für die Symmetrieelemente gibt es anschauliche geometrische Interpretationen.

1. *Drehachsen.* Eine Achse heißt n-zählige Drehachse, wenn der Kristall nach einer Drehung um $2\pi/n$ erstmals in sich selbst übergeführt wird (Abb. 6.1.3–1). Als Folge der zusätzlich vorhandenen Translationssymmetrie sind nur die Zähligkeiten $n = 1, 2, 3, 4$ und 6 möglich.

2. *Drehinversionsachsen.* Eine Achse heißt n-zählige Drehinversionsachse, wenn der Kristall nach einer Inversion, d. h. Spiegelung am Koordinatensprung, und anschließender Drehung um $2\pi/n$ mit sich selbst zur Deckung kommt (Abb. 6.1.3–2). Die Zähligkeiten n sind die gleichen wie bei den Drehachsen. Eine Drehinversionsachse der Zähligkeit 2 wird auch als *Spiegel-* oder *Symmetrieebene* bezeichnet.

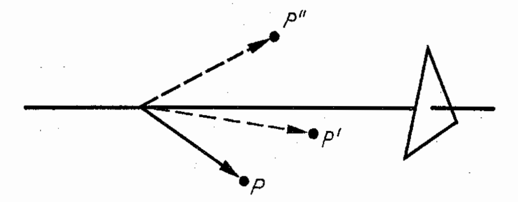


Abb. 6.1.3–1 Dreizählige Drehachse, die die Punkte P, P' und P'' verknüpft (grafisches Symboldreieck). Nach Drehung um $2\pi/3$ decken sich P mit P', P' mit P'' und P'' mit P

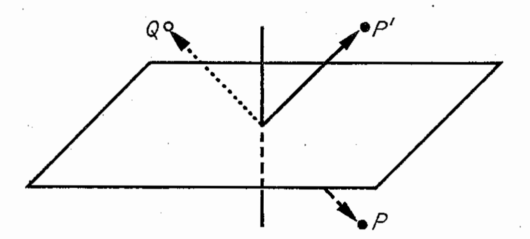


Abb. 6.1.3–2 Drehinversionsachse, die Punkte P und P' sind durch eine zweizählige Drehinversion miteinander verknüpft. Nach Inversion geht P in Q über und durch anschließende Drehung um $2\pi/2$ wird P' erreicht

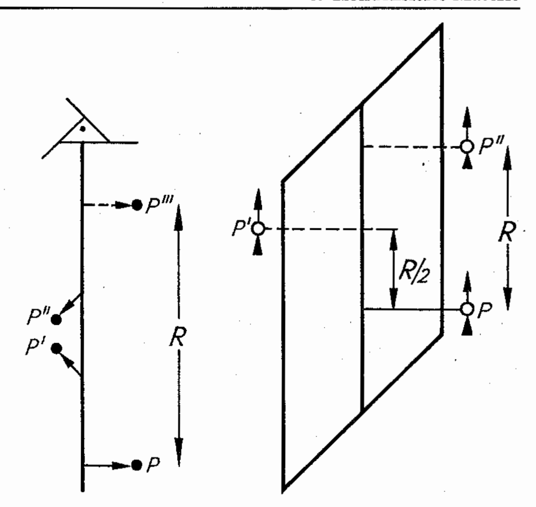


Abb. 6.1.3–3 (links) Die Punkte P, P', P'' und P''' sind durch eine dreizählige Schraubenachse mit $p = 1$ ineinander überführbar. P gelangt nach P' infolge Drehung um $2\pi/3$ und Verschiebung um $\boldsymbol{R}/3$

Abb. 6.1.3–4 Durch Spiegelung an einer Ebene und anschließender Verschiebung um $\boldsymbol{R}/2$ gelangt P nach P' und P' nach P'' (Gleitspiegelung)

3. *Schraubenachsen.* Eine Schraubenachse liegt vor, wenn die Deckungsgleichheit nach Drehung um $2\pi/n$ und Verschiebung parallel zur Drehachse um $(p/n)\,\boldsymbol{R}$ mit $p = 0, 1, \ldots, (n-1)$ erreicht wird (Abb. 6.1.3–3). $\boldsymbol{R}$ ist der Gittervektor in Richtung der Drehachse. Auch hier kann n die Werte 1, 2, 3, 4, 6 annehmen.

4. *Gleitspiegelebenen.* Kann der Kristall durch Spiegelung aller Bestandteile an einer Ebene und anschließender Parallelverschiebung um $\boldsymbol{R}/2$ mit sich selbst zur Deckung gebracht werden, liegt eine Gleitspiegelebene vor (Abb. 6.1.3–4). $\boldsymbol{R}$ ist ein Gittervektor parallel zur Spiegelebene. Nach dessen Orientierung unterscheidet man fünf Gleitspiegelebenen.

Darüber hinaus können *Antisymmetrieoperationen* bzw. -elemente vorliegen, die eine Eigenschaftsänderung der betreffenden Bestandteile ausdrücken, z. B. Vorzeichenänderung der Ladung.

6.1.3.3. Kombinationen von Symmetrieoperationen. Gewöhnlich werden in Kristallen mehrere Symmetrieoperationen bzw. -elemente nebeneinander beobachtet. Die Zahl der Kombinationsmöglichkeiten unter Berücksichtigung der räumlichen Anordnung ist beschränkt. Sie beträgt für Drehachsen und Drehinversionsachsen 32. Dabei bleibt ein Punkt des Raumes fest. Da diese Kombinationen Gruppen im Sinne der Gruppentheorie bilden, spricht man von 32 *Punktgruppen* (*Kristallklassen*).

Wird die Translation in die Kombination einbezogen, werden also zusätzlich noch Schraubenachsen und Gleitspiegelebenen berücksichtigt, ergeben sich 230 mögliche Kombinationen, die *Raumgruppen*. Infolge der wirkenden Translation bleibt kein Punkt des Raumes fest, wenn der (als unendlich groß angesehene) Kristall mit sich selbst zur Deckung gebracht wird.

Die Berücksichtigung der Antisymmetrieoperation führt zu den 1651 *Heesch-Schubnikov-Gruppen*.

6.1.3.4. Kristallsysteme. Die konventionelle Wahl der Koordinatenachsen in den 14 Gittern (vgl. Abb. 6.1.2–1) erfolgt so, daß Drehachsen oder Drehinversionsachsen möglichst hoher Zähligkeit parallel zu ihnen liegen. Das führt auf sieben verschiedene Koordinatensysteme (vgl. die sieben Zeilen in Abb. 6.1.2–1). Alle Gitter, Punkt- oder Raumgruppen, die durch das gleiche Koordinatensystem beschrieben werden, gehören zum gleichen Kristallsystem.

6.1.4. Kristallstrukturen

Unter der Kristallstruktur eines Stoffes wird gewöhnlich die Gesamtheit der zeitlich gemittelten Koordinaten aller Atomkerne bezüglich der konventionellen Elementarzelle (vgl. Abb. 6.1.2–1) verstanden. Der Einfluß von Kristallbaufehlern wird nach Möglichkeit eliminiert. Die Koordinaten sind in umfangreichen Tabellenwerken und elektronischen Datenbanken zusammengestellt. Zeichnerische Darstellungen verwenden meist Kugeln unterschiedlichen Durchmessers für die Bestandteile (vgl. Abb. 6.1.1–1). In zunehmendem Maße werden auch Angaben zur räumlichen Verteilung der Elektronen und der magnetischen Momente zur Kennzeichnung der Kristallstruktur verwendet. Der Mannigfaltigkeit der Strukturen sind im Gegensatz zu ihren Gittern (vgl. 6.1.2.) keine Grenzen gesetzt. Man kann die Vielzahl der bis zum gegenwärtigen Zeitpunkt entdeckten Strukturen durch eine Reihe geeigneter Begriffsbildungen überschaubar machen, d. h. klassifizieren.

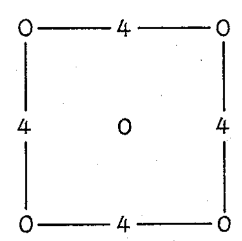


Abb. 6.1.4–1 F-Komplex als Beispiel für einen Gitterkomplex. Die Ziffern geben die Koordinate der Punkte senkrecht zur Bildebene in Achteln der Gitterkonstante an, die durch die Länge einer Quadratseite gegeben ist

6.1.4.1. Gitterkomplex. Die Gesamtheit aller Positionen in der Elementarzelle, die durch Anwendung der Raumgruppensymmetrieoperationen auf einen Punkt entstehen, etwa im Sinne der Abb. 6.1.3–1 bis –4, nennt man Gitterkomplex. An Stelle der Angabe unübersichtlicher Punktkoordinaten können Hinweise auf die von gewissen Atomen besetzten geometrisch durchsichtigen Gitterkomplexe treten und das Verständnis der Struktur fördern. Abb. 6.1.4–1 zeigt z. B. den sogenannten F-Komplex. Die NaCl-Struktur (vgl. Abb. 6.1.1–1) läßt sich beschreiben als F-Komplex jeweils für Na- und Cl-Atome.

6.1.4.2. Atomgrößen. Die Atomkerne sind von Elektronen umgeben, die sich, wie Messungen der Elektronendichte zeigen, weit in die Elementarzelle hinein erstrecken. Der bloße Befund des Auftretens bestimmter polyedrischer Atomanordnungen führt zu der Vorstellung von Atom„größen“. Im einfachsten Falle wird das Atom durch eine Kugel mit dem Atomradius r beschrieben, oder das Ion durch den Ionenradius. Trotz der Abhängigkeit des Radius vom Bindungstyp (r[Mg] $= 1{,}37 \cdot 10^{-10}$ m für homöopolare, $1{,}6 \cdot 10^{-10}$ m für metallische und $0{,}75 \cdot 10^{-10}$ m für heteropolare Bindung) wird auch heute noch nach verbesserten Werten gesucht.

6.1.4.3. Kugelpackungen. Wird den Bausteinen der Kri-

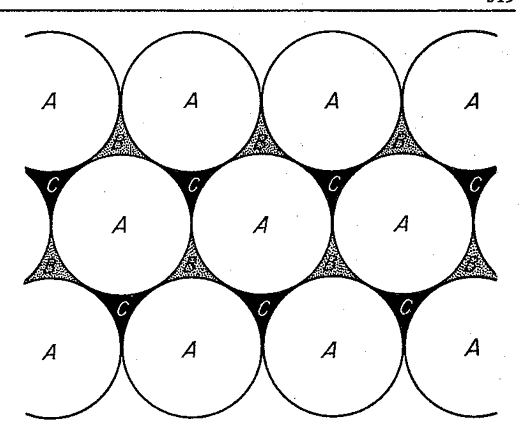


Abb. 6.1.4–2 Dichte Kugelpackung. Die beiden verschiedenen Möglichkeiten, eine zweite derartige Schicht in stabiler Lage auf die erste zu setzen, sind durch B und C angedeutet

stallstruktur Kugelgestalt zugeschrieben, gelangt man dazu, die zahlreichen unterschiedlichen Strukturen als Varianten von Kugelpackungen zu betrachten. Jede Kugel berührt nach dieser Vorstellung mindestens eine andere. Die Kugeln können gleiche oder unterschiedliche Radien aufweisen. In zeichnerischen Strukturdarstellungen (vgl. Abb. 6.1.1–1) werden die Kugeln aus Gründen der besseren Durchschaubarkeit meist mit reduziertem Radius dargestellt.

Zahlreiche Kristallstrukturen erweisen sich auf diese Weise als geometrisch nahe verwandt, sind sie doch aus identischen Stapelschichten lediglich mit unterschiedlicher *Stapelfolge* aufgebaut (Abb. 6.1.4–2). So läßt sich z. B. Magnesium durch eine Schichtfolge ABAB..., Kupfer durch ABCABC... und Praseodym durch ABACABAC... beschreiben.

6.1.4.4. Strukturtypen. Strukturen mit gleichen Gitterkomplexen (vgl. 6.1.4.2.) und der gleichen Raumgruppe (vgl. 6.1.3.3.) werden zu einem Strukturtyp zusammengefaßt. Er vereinigt damit Anordnungen von Atomen gleicher Symmetrie und gleicher Zahl in der Elementarzelle ohne Rück-

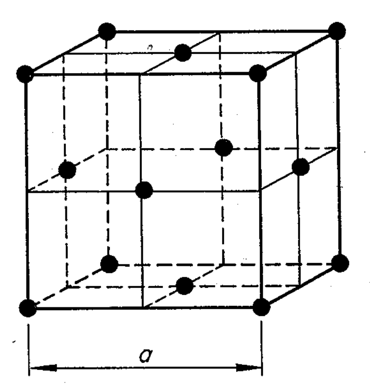


Abb. 6.1.4–3 Konventionelle Elementarzelle des Cu-Typs. Die besetzten Atomlagen sind mit den Gitterpunkten des kubisch-flächenzentrierten Gitters (Gitterkonstante a) identisch

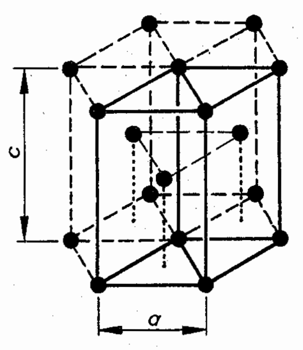


Abb. 6.1.4–4 Elementarzelle des Mg-Typs (durchgehende Linien). Das Gitter ist hexagonal (Gitterkonstanten *c* und *a*). Mit jedem Gitterpunkt sind zwei Atome verknüpft (schwarze Kreise). Um die Sechszähligkeit besser hervortreten zu lassen, ist mehr als eine Elementarzelle dargestellt

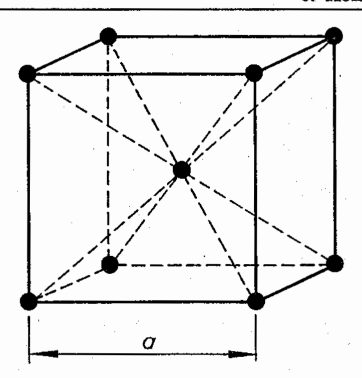
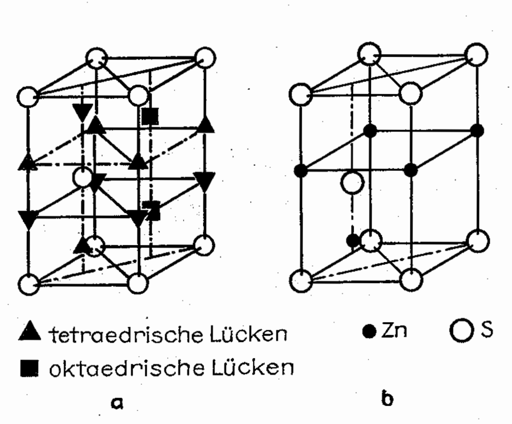


Abb. 6.1.4–5 Konventionelle Elementarzelle des W-Typs. Die besetzten Atomlagen sind mit den Gitterpunkten des kubisch-raumzentrierten Gitters (Gitterkonstante *a*) identisch



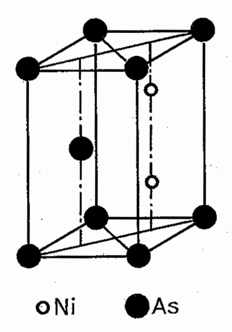


Abb. 6.1.4–7 Aufbau der Strukturtypen des Wurtzit (ZnS) (b) und des NiAs (c) durch Füllung der Hälfte der tetraedrischen bzw. aller oktaedrischen Lücken des Mg-Typs (a)

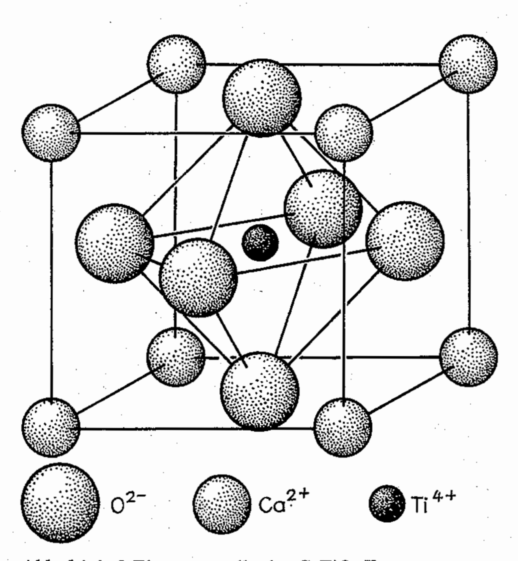


Abb. 6.1.4–8 Elementarzelle des $CaTiO_3$-Typs

sicht auf die absolute Größe der Zelle und der Parameter von Atomlagen. Substanzen des gleichen Strukturtypes weisen oft Gemeinsamkeiten in der Bindung auf.

Strukturtypen werden heute überwiegend nach einem Vertreter (Prototyp) benannt, z. B. kristallisieren die Elemente Germanium, Silicium und Zinn im *Diamanttyp*, d. h. mit dem gleichen Strukturtyp wie Diamant.

Werden Ionen oder Atome entsprechend 6.1.4.3. durch Kugeln dargestellt, bleiben *in der Elementarzelle Lücken unterschiedlicher Größe*. Deren Anordnung ist für die Interpretation physikalischer Eigenschaften, z. B. der Diffusion, und zur Diskussion von geometrischen Verwandtschaftsbeziehungen zwischen unterschiedlichen Strukturtypen von Bedeutung. Im folgenden werden einige häufig auftretende Strukturtypen erläutert.

Die *metallischen Elemente* einschließlich aller zugehörigen Mischkristalle und eine Reihe Nichtmetalle, z. B. feste Edelgase, kristallisieren in der überwiegenden Zahl in einem der drei nachstehenden Strukturtypen mit zwölf bzw. acht nächsten Nachbarn eines jeden Atoms (Abb. 6.1.4–3 bis –5).

Cu-Typ: Ac, Ag, Al, Am, Ar, Au, Ca, Ce, β-Co, Cu, γ-Fe, Ir, Kr, β-La, Li, γ-Mn, Ne, Ni, Pb, Pd, Pt, δ-Pu, Rh, α-Sr, α-Th, Xe, α-Yb.

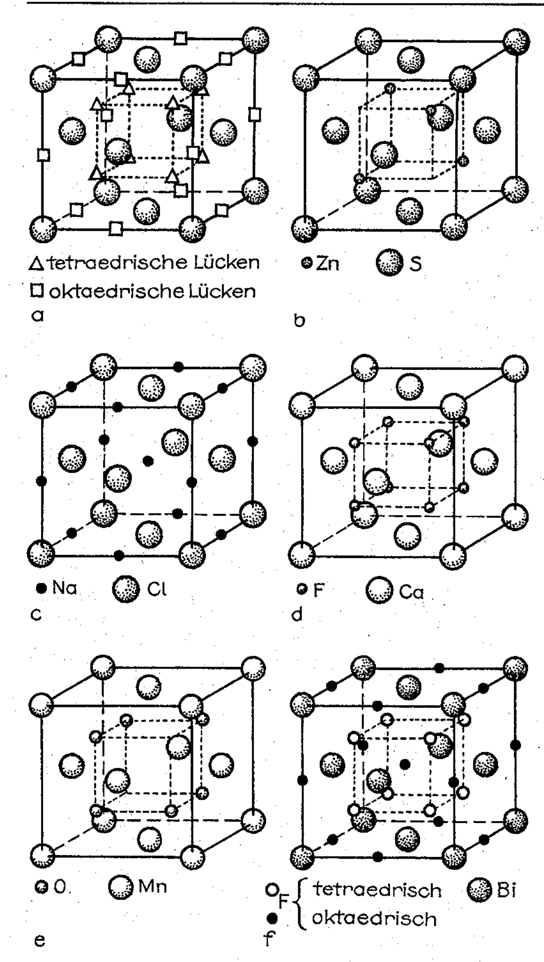


Abb. 6.1.4–6 Aufbau der Strukturtypen Diamant und Zinkblende (ZnS) (b), NaCl (c), CaF_2 (d), Mn_2O_3 (e) und BiF_3 (f) durch unterschiedliche Füllung der tetraedrischen und oktaedrischen Lücken des Cu-Typs (a). Im Diamanttyp sind Zn- und S-Lagen von C besetzt. Bei Mn_2O_3 sind 3/4 der Lücken besetzt, so daß sich die vollständige Elementarzelle aus acht Würfeln der dargestellten Form zusammensetzt

Mg-Typ: Be, Cd, β-Ce, Co, Dy, H, Hf, Ho, La, β-Li, Lu, Mg, Na, α-Nd, Os, Pm, Pr, Re, Ru, Sc, β-Sr, Tb, Tc, Ti, Tl, Tm, Y, Zn, Zr.

W-Typ: Ba, β-Be, γ-Ca, δ-Ce, Cr, Cs, Eu, α-Fe, δ-Fe, β-Gd, β-Hf, K, γ-La, Li, δ-Mn, Mo, Na, Nb, δ-Nd, γ-Np, β-Pr, ε-Pu, Rb, β-Sm, γ-Sr, Ta, β-Th, β-Ti, β-Tl, γ-U, V, W, β-Y, β-Yb, β-Zr.

Im schon erwähnten Diamanttyp dominiert die tetraedrische Koordination. Der Diamanttyp und einige weitere wichtige Strukturtypen lassen sich auch durch unterschiedliche Füllung der Lücken aus dem Cu-Typ ableiten. Aus Abb. 6.1.4–6 geht dies für Diamant-, Zinkblende-, NaCl-, CaF_2- Mn_2O_3- und BiF_3-Typ unmittelbar hervor. Analog folgen aus dem Mg-Typ die Strukturtypen von Wurtzit (ZnS) und NiAs (Abb. 6.1.4–7).

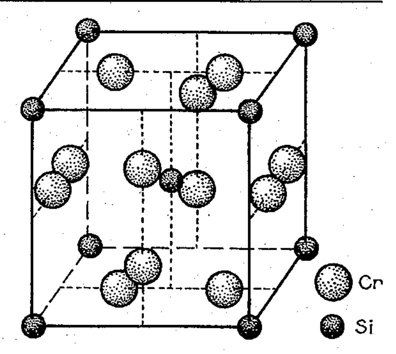


Abb. 6.1.4–9 Elementarzelle des Cr_3Si-Typs. Entlang der gestrichelten Linien liegen die Cr-Atome im Kristall besonders dicht beieinander

Weitere Beispiele für aus dichtesten Kugelpackungen abgeleiteten Strukturtypen stellen $MgCu_2$, $MgZn_2$ und $MgNi_2$ (Lavesphasen) dar. Füllt man im Cu-Typ die Hälfte der tetraedrischen Lücken mit Mg-Atomen, den Rest mit Cu-Tetraedern, erhält man den $MgCu_2$-Typ.

Der wegen seiner ferroelektrischen Vertreter bekannte $CaTiO_3$-Typ (Abb. 6.1.4–8) ist insofern ein besonderer Abkömmling der dichtesten Kugelpackung, als hier Anionen und größere Kationen gemeinsam den F-Komplex besetzen und in einer oktaedrischen Lücke das kleinere Kation untergebracht ist. Beim Cr_3Si-Typ (Abb. 6.1.4–9) ist die Hälfte der tetraedrischen Lücken des W-Typs durch Cr-Atome besetzt. Dadurch entsteht ein System windschiefer Cr-Ketten. Einige hervorragende Supraleiter kristallisieren in diesem Typ. Neben Ersetzungsmischkristallen (Austausch von Atomen) sind auch Einlagerungsmischkristalle möglich. Sie entstehen durch ungeordnete Lückenfüllung.

Schon bei den Elementen beobachtet man bei Variation von Druck, Temperatur u. a. Parametern verschiedene Strukturtypen an chemisch gleichen Substanzen (Polymorphismus). Das trifft in besonderem Maße auf Legierungen oder Verbindungen zu.

6.2. Herstellung von Kristallen

6.2.1. Kristallkeimbildung

Kristallkeime können sich nur bilden, wenn das System durch Übersättigung (Unterkühlung) aus dem thermodynamischen Gleichgewicht gebracht wird. Treffen dann genügend Teilchen infolge statistischer Schwankungen der Dichte und Geschwindigkeit gleichzeitig zusammen, so erfolgt eine *spontane Keimbildung*, die homogen ist. Lagern sich die Teilchen an Fremdpartikeln oder Gefäßwände an, ist die Keimbildung heterogen. Die Anzahl der je Zeit- und Volumeneinheit bei konstanter Übersättigung gebildeten stabilen Keime, die die *Kristallisationsgeschwindigkeit* bestimmt, ist die *Keimbildungshäufigkeit*. Sie hängt ab von der *Diffusion* der Teilchen zum Keimort und von der *Keimbildungsarbeit*. Letztere muß aufgebracht werden, um den Keim aus den Teilchen aufzubauen. Sie ist die Differenz zwischen zwei Energiebeträgen. Ein Anteil, der der Übersättigung und der Anzahl der Teilchen proportional ist, wird bei der Zusammenlagerung der Teilchen frei, da die ungeordnete Phase stets eine höhere freie Enthalpie besitzt

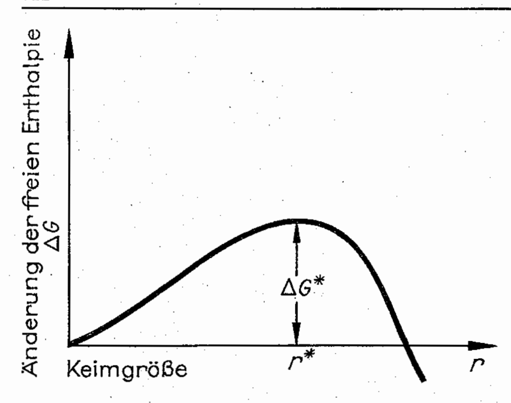


Abb. 6.2.1–1 Abhängigkeit der Gesamtänderung der freien Enthalpie von der Keimgröße bei konstanter Unterkühlung

als die kristalline. Der zweite Anteil ist notwendig, um die *freie Oberflächenenthalpie* des Keimes aufzubringen und ist der gebildeten Oberfläche des Keimes proportional. Durch das Gegeneinanderwirken beider Beträge durchläuft die Gesamtänderung der freien Enthalpie ΔG in Abhängigkeit von der Keimgröße ein Maximum (Abb. 6.2.1–1). Im aufsteigenden Ast überwiegt der Einfluß der freien Oberflächeenthalpie. Eine Zunahme der Keimgröße führt zur Erhöhung von ΔG. Der Keim ist nicht stabil und zerfällt. Im Maximum wird die *kritische Keimbildungsarbeit* ΔG^* für die Bildung stabiler Keime erreicht. Eine weitere Anlagerung von Teilchen an einen Keim der kritischen Größe r^* ist mit einem Abfall von ΔG verbunden und führt zur Kristallisation. Während r^* und ΔG^* mit steigender Unterkühlung abnehmen, streben sie bei sehr kleiner Unterkühlung unendlich großen Werten zu. Die theoretische Berechnung von ΔG^* und r^* ist für die Kondensation von Metallen aus der Gasphase unter Ultrahochvakuumbedingungen näherungsweise möglich. Der Keimbildungsprozeß kann durch Fremdatome, Erschütterungen, Licht, energiereiche Strahlung, elektrische und magnetische Felder beeinflußt werden. Besonders unübersichtlich verläuft er in Schmelzen und Lösungen.

In Schmelzen durchläuft die *Keimbildungsgeschwindigkeit* mit steigender Unterkühlung ein Maximum, da die Zunahme der Viskosität der Schmelze die Diffusion der Teilchen zum Keimort hemmt.

Die *heterogene Keimbildung* kann auf amorphen und kristallinen Stoffen (*Substraten*) erfolgen. Ähnlichkeiten in Kristallstruktur und chemischer Bindung beider Partner führen oft zu einer intensiven Wechselbeziehung, so daß die Keimbildungsarbeit im Vergleich zur homogenen Keimbildung stark erniedrigt wird. Es können *orientierte*, *gesetzmäßige Verwachsungen von unterschiedlichen Kristallen* entstehen (*Epitaxie*). Wahrscheinlich ist auch Silberiodid für die Wasserdampfkondensation deshalb besonders wirksam, weil es mit Eis isomorph ist (Hagelbekämpfung).

6.2.2. Kristallwachstum

6.2.2.1. Molekularkinetische Theorie. Sie geht davon aus, daß die aus der ungeordneten Phase auf der Kristallfläche adsorbierten Teilchen durch *Oberflächendiffusion* zu freien Gitterstellen wandern, an denen sie mit der größten Anlagerungsenergie gebunden werden. Es werden zunächst Ketten aufgefüllt und begonnene Netzebenen vervollständigt. Auf einer atomar glatten Fläche muß erst ein stabiler *zweidimensionaler Netzebenenkeim* gebildet werden, ehe eine erneute Anlagerung unter Energiegewinn erfolgen kann. Die für diesen Vorgang notwendige Unterkühlung ist geringer als bei der homogenen Keimbildung. Bei der Weiterentwicklung dieser Theorie wurden Faktoren berücksichtigt, die das reale Kristallwachstum stark beeinflussen, so z. B. der simultane Einbau von Teilchen und die Bildung neuer Netzebenenkeime sowie die Beeinflussung der Keimbildungs- und Wachstumsgeschwindigkeit durch Adsorption von Fremdteilchen. Es ist experimentell nachgewiesen, daß *Gitterversetzungen* (vgl. 6.5.2.) die Bildungsarbeit herabsetzen können. Der Wachstumsmechanismus über *Schraubenversetzungen* verläuft bei besonders geringer Unterkühlung, da die Bildungsarbeit für einen Netzebenenkeim entfällt.

6.2.2.2. Wachstums- und Gleichgewichtsformen. Makroskopische Kristalle wachsen niemals in allen Richtungen gleichmäßig. Die *linearen Wachstumsgeschwindigkeiten* senkrecht zu den Kristallflächen sind unterschiedlich, wie ein Kugelwachstumsversuch zeigt. Dazu wird eine durch mechanische Bearbeitung hergestellte Kristallkugel in eine übersättigte Lösung gebracht. Beim Wachstumsprozeß verschwinden energetisch ungünstige Flächen mit hoher line-

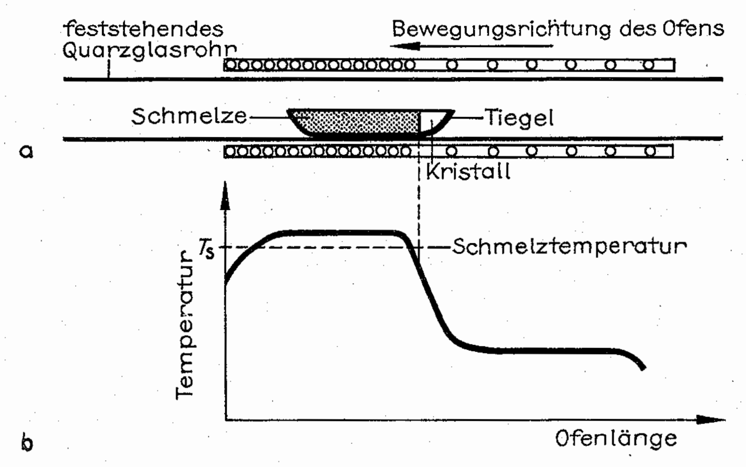


Abb. 6.2.4–1 (a) Prinzip der Bridgman-Züchtung, (b) Temperaturprofil (T_s = Schmelztemperatur)

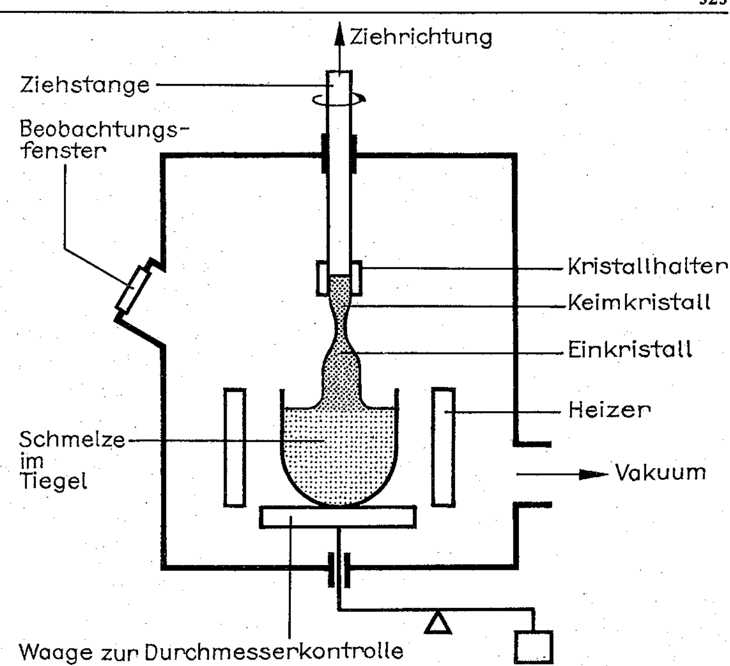


Abb. 6.2.4–2 Czochralski-Verfahren

arer Wachstumsgeschwindigkeit, und es bilden sich *stationäre Wachstumsformen* heraus. Sie können durch Gefäßwände oder gegenseitige Berührung von Kristallen gestört werden. Unter speziellen Bedingungen können stark verzerrte Formen, wie *Dendriten* (skelettartige verzweigte Kristalle) oder *Whiskers* (dünne nadelförmige Haarkristalle), auftreten.
Abweichend davon ist die *Gleichgewichtsform*, die bei gegebenen Volumen die *geringste freie Oberflächenenthalpie* aufweist. Sie wird durch Flächen kleinster linearer Wachstumsgeschwindigkeit begrenzt, die mit den Flächen kleinster Oberflächenenthalpie identisch sind. Die Gleichgewichtsform ist praktisch nicht realisierbar, da der Kristall während seines Wachstums mit der umgebenden Phase nicht im Gleichgewicht steht. An Wachstumsformen können jedoch Gleichgewichtsflächen auftreten.

6.2.3. Einkristalle und Polykristalle

Einzelne große Kristalle (*Einkristalle*) entstehen bei *geringer Keimbildungsgeschwindigkeit* (vgl. 6.2.1.) oder durch Vorgabe eines makroskopischen *Keimkristalles*. Sie sind für die Festkörperphysik von besonderem Interesse, da an ihnen mit der Gitterperiodizität und Anisotropie verbundene Eigenschaften studiert werden können. *Polykristalle* entstehen bei hoher Keimbildungsgeschwindigkeit durch *spontane Kristallisation* entweder als *Pulver* oder *Haufwerk* verschiedener Kristallindividuen oder als kompaktes *Gefüge* mit Kristalliten verschiedener Orientierung wie bei Gußeisen.

6.2.4. Einkristallzüchtung

6.2.4.1. Erstarrung aus Schmelzen. Sie erfordert eine Unterschreitung der Schmelztemperatur, wobei die Erstarrungsgeschwindigkeit von der Abfuhr der freiwerdenden *Kristallisationswärme* abhängt. Für die gerichtete Erstarrung einer Schmelze gibt es verschiedene experimentelle Anordnungen.
Beim *Bridgman-Verfahren* erfolgt eine Relativbewegung der Schmelze zum Temperaturfeld eines Mehrzonenofens (Abb. 6.2.4–1). Für die vertikale Variante werden oft Behälter verwendet, bei denen der Teil, in dem die Kristallisation beginnt, eine spezielle Geometrie hat. So wird durch eine Spitze die Einkristallbildung gefördert, da nur wenige Kristallite parallel zur Tiegelachse weiterwachsen können.
Beim *Čochralski-Verfahren* wird ein in eine Schmelze eingetauchter Keimkristall bei gleichzeitiger Rotation mit einer Geschwindigkeit bis zu 100 cm/h nach oben gezogen (Abb. 6.2.4–2, Taf. 45). Der Kristalldurchmesser hängt von der Ziehgeschwindigkeit ab. Sie wird bei modernen Anlagen z. B. durch die Messung der Gewichtsänderung der Schmelze gesteuert. Störungen im gewachsenen Kristall können durch Verwendung perfekter einkristalliner Keime, Erhöhung der Ziehgeschwindigkeit in der Anfangsphase (*Dünnhalstechnik*) und Nachheizer zur Verringerung des Temperaturgradienten nahe der Grenzfläche fest – flüssig reduziert werden. Das Verfahren ist auch für Substanzen anwendbar, die bei hohem Druck unter Zersetzung schmelzen. Bei Galliumphosphid ($T_s = 1740$ K, $p = 3{,}55$ MPa) wird eine transparente Bor(III)-oxid-Abdeckschmelze benutzt und die Phosphorverdampfung durch Kompensation mit einem angemessenen Inertgasdruck verhindert. Beim *Kyropoulos-Verfahren* wird der Keimkristall während der Züchtung erzeugt, indem der doppelwandige rohrartig ausgebildete Kristallhalter mit einem steuerbaren Gasstrom gekühlt wird (Abb. 6.2.4–3). Durch geringfügiges Anheben des Halters zu Beginn des Züchtungsprozesses findet die *Keimauswahl* statt. Das weitere Wachstum wird durch den Kühlgasstrom gesteuert. Die untere Begrenzungsfläche des Kristalles ist identisch mit dem Verlauf der Schmelzisother-

men im Züchtungstiegel. Sie verleiht dem Kristall die charakteristische Kalottenform (Taf. 45).

Beim *Zonenschmelzen* wird nur ein schmaler Bereich eines polykristallinen Barrens aufgeschmolzen. Durch Bewegung des Heizers wandert diese Zone durch ihn hindurch. Ursprünglich wurde das Verfahren zur *Hochreinigung* von Metallen benutzt. Das Prinzip beruht auf der unterschiedlichen Löslichkeit von Fremdstoffen in der flüssigen (c_l) und der festen Phase (c_s). Die Reinigungswirkung ist um so

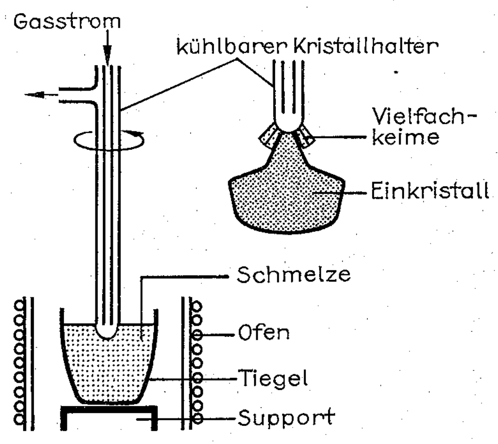


Abb. 6.2.4–3 Prinzip der Kyropoulos-Anlage

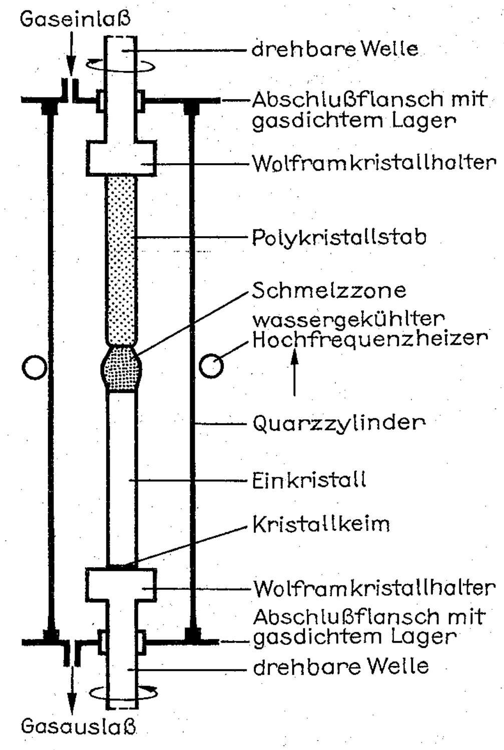


Abb. 6.2.4–4 Tiegelfreies Zonenschmelzen

größer, je mehr sich der *Gleichgewichtsverteilungskoeffizient* $k_0 = c_s/c_l$ einer Fremdkomponente von Eins unterscheidet. Der *effektive Verteilungskoeffizient* hängt von k_0, der Erstarrungsgeschwindigkeit der Schmelze und der Diffussion der Fremdkomponente in der flüssigen Phase ab. Mehrere Zonendurchgänge in einer Richtung verstärken die Trennung der Komponenten. Eine *homogene Verteilung* wird erreicht, wenn die Fremdkomponente aus der Gasphase in das System eingebracht und die Schmelzzone in beiden Richtungen bewegt wird (*zone leveling technique*). Dieses Verfahren wird zur Dotierung von Halbleitern angewandt.

Um den Einbau von Verunreinigungen aus dem Tiegelmaterial zu verhindern und Wandeffekte auszuschalten, wurde das *tiegelfreie Zonenschmelzen* (*floating zone technique*) entwickelt (Abb. 6.2.4–4). Von einem kompakten Stab wird eine schmale Zone aufgeschmolzen und durch den Stab bewegt. Man verwendet dazu Hochfrequenz- bzw. optische Heizer, bei denen das Licht einer Halogenlampe durch Parabolspiegel stark gebündelt wird. Beide Teile des durch die Schmelzzone getrennten Stabes rotieren gegenläufig. Ein kritischer Parameter ist die maximale Höhe der Schmelzzone. Sie hängt von der Oberflächenspannung und Dichte der Schmelze ab. Daher ist das Verfahren auf bestimmte Stoffe beschränkt. Gegenwärtig werden damit hochreine versetzungsfreie Siliciumkristalle von 50 bis 80 cm Länge, 10 bis 15 cm Durchmesser und einer Fremdstoffkonzentration von 10^{12} Atomen cm^3 gezüchtet.

Das *Verneuil-Verfahren* verwendet feingepulvertes Ausgangsmaterial, das in einer heißen Gasflamme auf einen Keim als dünner Film aufgeschmolzen wird. Es entsteht ein birnenförmiger Einkristall. Mit dieser Technik werden industriell synthetische Edelsteine hergestellt.

6.2.4.2. Kristallisation aus Lösungen. Sie verläuft unterhalb des Schmelzpunktes und ist für Substanzen mit hohen Schmelztemperaturen, Zersetzung am Schmelzpunkt oder Phasenumwandlungen im festen Zustand wichtig. Die Kenntnis der Löslichkeit in Abhängigkeit von der Temperatur entscheidet darüber, ob die Übersättigung durch Abkühlung der Lösung oder isotherme Verdampfung des Lösungsmittels erzeugt wird. Abkühlung erfolgt bei starker Temperaturabhängigkeit, isotherme Verdampfung bei schwacher Temperaturabhängigkeit der Löslichkeit. Die Kristallisationsgeschwindigkeit wird durch die Diffusion bestimmt. Sie liegt bei 10^{-5} cm/h.

Technisch kann die Kristallisation durch Abkühlung einer Lösung z. B. in einem *Drei-Tank-Verfahren* realisiert werden (Abb. 6.2.4–5). Hierbei wird die Lösung im Tank I gesättigt. Zur Vermeidung von *spontaner Keimbildung* wird sie im Tank II einige Grad überhitzt und mit einer Pumpe über eine wärmeisolierte Leitung in den Tank III befördert, dessen Temperatur niedriger als die im Tank I ist. Durch das Kristallwachstum verarmt die Lösung an Substanz und wird im Tank I wieder gesättigt und der Kreislauf beginnt von neuem. Damit wurden Ammoniumdihydrogenphosphatkristalle von etwa 1 kg aus wäßriger Lösung gezüchtet.

Die Lösungskristallisation ist nicht auf Niedertemperaturlösungen beschränkt. Oxidische Materialien können bei hohen Temperaturen aus Mischungen von Blei(IV)-oxid, Blei(II)-fluorid und Bor(III)-oxid, Verbindungshalbleiter aus Metallen kristallisiert werden.

Das *Hydrothermalverfahren*, bei dem die notwendige Löslichkeit durch gleichzeitige Temperatur- und Druckerhöhung erreicht wird, hat sich bei der Züchtung von piezoelektrischen α-Quarzkristallen aus alkalischen Silicatlösungen bewährt. Dazu werden Spezialautoklaven verwendet.

6.2.4.3. Kristallisation aus der Gasphase. Einige organische und anorganische Verbindungen besitzen bereits unterhalb ihres Schmelzpunktes einen hohen Dampfdruck. Sie können von hoher Temperatur (T_2) zu tiefer Temperatur (T_1) sublimieren. Einfache Versuchsanordnungen sind Glasampullen oder ein offenes Rohr, in dem ein Inertgasstrom über die Probe geleitet wird.
Festkörper mit geringem Dampfdruck können mit einer Zusatzkomponente B zu einer gasförmigen Verbindung C reagieren. $i A_{\text{fest}} + j B_{\text{gas}} \rightleftarrows k C_{\text{gas}}$ (i, j, k = Stöchiometriekoeffizienten). Ist die Enthalpie dieser Gleichgewichtsreaktion endotherm, so wird A bei T_2 in der Gasphase gelöst und scheidet sich bei T_1 wieder ab. Der umgekehrte Fall tritt ein, wenn sie exotherm ist (*chemischer Transport*). Die im Wachstumsraum notwendige Übersättigung ergibt sich aus dem Unterschied der Partialdrücke der Verbindung C bei den Temperaturen T_2 und T_1. Der Massetransport vom Lösungsraum zum Wachstumsraum erfolgt bei Gesamtdrükken von 10 bis 100 kPa durch *Diffusion*, bei höheren Drükken durch *Konvektion*. Das Transportmittel darf nicht in den Festkörper eingebaut werden, da der Prozeß sonst zum Stillstand kommt. Wegen ihrer Reaktionsfähigkeit und ihrer großen Atomvolumina haben sich die Halogene bzw. Halogenwasserstoffe besonders bewährt. Es können Kristalle von mehreren Zentimetern Länge von Verbindungen mit hohem Schmelzpunkt gezüchtet und *epitaktische Schichten* (vgl. 6.2.1.) auf einkristallinen Substraten abgeschieden werden (*chemical vapor phase deposition*). In der Halbleitertechnik dienen letztere zur Herstellung von integrierten Schaltkreisen oder optoelektronischen Bauelementen.
Andere Verfahren zur Schichtabscheidung sind die explosionsartige Verdampfung von Substanzen bei hohen Temperaturen (*Flash-Technik*) oder das Zerstäuben von Festkörpern durch Beschuß mit *Elektronen-* oder *Ionenstrahlen* im Vakuum.

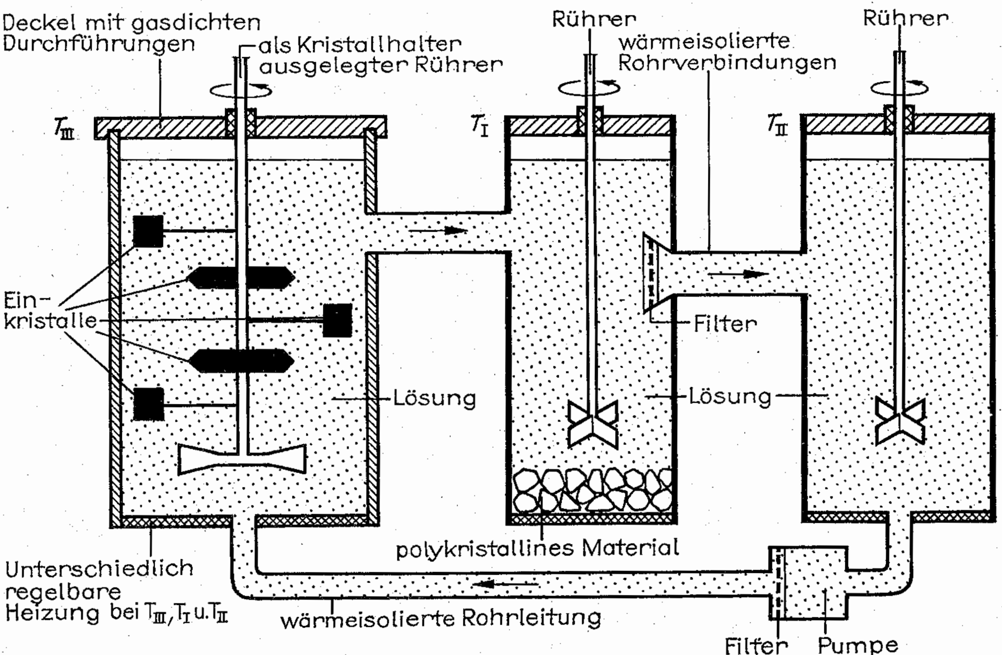


Abb. 6.2.4–5 Drei-Tank-Kristallisator, Temperaturen der Tanks $T_{III} < T_I < T_{II}$

6.3. Elektronenzustände

Für ein einzelnes Atom führte die Kenntnis der Energieniveaus der Elektronen und der zugehörigen Wellenfunktionen zu einem anschaulichen Bild, aus dem man die Atomspektren u.a. atomare Eigenschaften verstehen kann. Daher wird man auch beim Festkörper von diesen Größen ausgehen, um aus mikroskopischer Sicht zu einem Verständnis seiner Eigenschaften zu gelangen. Wegen der größeren Komplexität eines Systems von größenordnungsmäßig 10^{23} Atomen wird dies nur in Ausnahmefällen mit der Klarheit gelingen, die beim einzelnen Atom erreicht worden ist. Bei einem *geordneten, kristallinen Aufbau* erleichtert die Ausnutzung der hohen Symmetrie, dieses Anliegen zu realisieren.

6.3.1. Translationssymmetrie

Die Atomlagen in einem geordneten Festkörper lassen sich von einer *Elementarzelle* ausgehend verstehen. Den gesamten Aufbau erhält man, indem man eine Elementarzelle um einen Gittervektor $R = l_1\boldsymbol{a}_1 + l_2\boldsymbol{a}_2 + l_3\boldsymbol{a}_3$ verschiebt, d.h. um ganzzahlige Vielfache der Basisvektoren $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ der Elementarzelle. Deren Volumen ist $\Omega_{\text{EZ}} = \boldsymbol{a}_1(\boldsymbol{a}_2 \times \boldsymbol{a}_3)$. Die reinen Metalle kristallisieren vorwiegend in den Strukturtypen *kubisch-raumzentriert* (krz), *kubisch-flächenzentriert* (kfz) und *hexagonal-dichtgepackt* (hdp) (vgl. 6.1.4.7.), für die hier die Basisvektoren

krz: $\boldsymbol{a}_1 = (-1, 1, 1)\, a/2$, $\boldsymbol{a}_2 = (1, -1, 1)\, a/2$,
$\boldsymbol{a}_3 = (1, 1, -1)\, a/2$,
kfz: $\boldsymbol{a}_1 = (0, 1, 1)\, a/2$, $\boldsymbol{a}_2 = (1, 0, 1)\, a/2$, $\boldsymbol{a}_3 = (1, 1, 0)\, a/2$,
hdp: $\boldsymbol{a}_1 = (\sqrt{3}/2, -1/2, 0)\, a$, $\boldsymbol{a}_2 = (0, 1, 0)\, a$,
$\boldsymbol{a}_3 = (0, 0, 1)\, c$,

gewählt werden. Ω_{EZ} hat die Werte $a^3/2$, $a^3/4$ bzw. $3a^2c/2$, ist beim krz- bzw. kfz-Typ also nur die Hälfte bzw. ein Viertel des Würfelvolumens von Abb. 6.1.4–5 bzw. 6.1.4–3.
Jeder reale Kristall hat eine *Oberfläche* mit spezifischen Eigenschaften (vgl. 6.10.). Die sogenannten Volumeneigenschaften sind unabhängig von der Größe der Probe und von der Form ihrer Oberfläche, d. h., sie sind unabhängig von den Randbedingungen an der Oberfläche. Zur Berechnung der Volumeneigenschaften wählt man daher solche Randbedingungen, die die Rechnung möglichst einfach gestalten. Dies sind sogenannte *periodische Randbedingungen*. Man unterteilt einen Festkörper in Gebiete mit je N Elementarzellen. Es wird nun vorausgesetzt, daß sich das physikalische Geschehen einschließlich der Randbedingungen

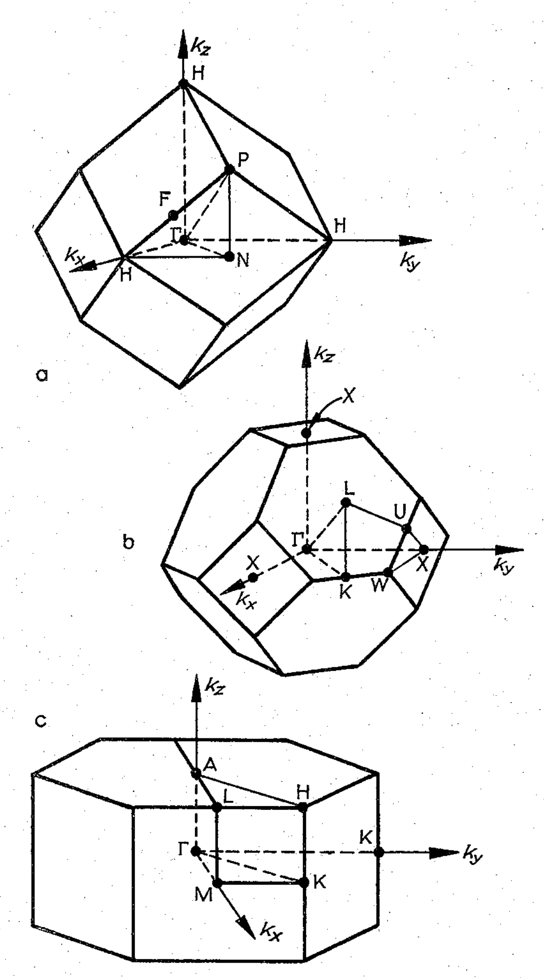


Abb. 6.3.1–1 Brillouin-Zone des (a) krz-, (b) kfz- und (c) hdp-Gitters

in den einzelnen Gebieten periodisch wiederholt. Dann müssen sich alle Größen am rechten Rand des herausgegriffenen Gebietes stetig an deren Werte am linken Rand anschließen usw. Im Eindimensionalen ist dies z. B. realisierbar, wenn man eine Kette von Atomen mit zwei Enden zu einem Ring zusammenschließt. In diesem Bild geht eine Atomanordnung beim Verschieben aller Atome um einen Gittervektor $\boldsymbol{R}$ in sich selbst über, während im realen Fall an der Oberfläche Veränderungen auftreten.
Die *Translationssymmetrie* beinhaltet die Eigenschaft, daß ein geordneter Festkörper beim Verschieben um einen Gittervektor $\boldsymbol{R}$ in sich selbst übergeht. Beispielsweise ist das Kristallpotential eine periodische Funktion $V(\boldsymbol{r}+\boldsymbol{R}) = V(\boldsymbol{r})$. Diese Translationssymmetrie weist auch der Hamilton-Operator auf, so daß für die Eigenfunktionen der Schrödinger-Gleichung $H\psi_k = E_k\psi_k$ das *Bloch-Theorem*

$$\psi_k(\boldsymbol{r}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}} u_k(\boldsymbol{r}),\; u(\boldsymbol{r}+\boldsymbol{R}) = u(\boldsymbol{r}) = \sum_{\boldsymbol{G}} c_G(\boldsymbol{k})\, \mathrm{e}^{\mathrm{i}\boldsymbol{G}\cdot\boldsymbol{r}}$$

folgt. Die Wellenfunktion unterscheidet sich in verschiedenen Elementarzellen nur durch einen Phasenfaktor $\psi_k(\boldsymbol{r}+\boldsymbol{R}) = \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}}\psi_k(\boldsymbol{r})$. Die Ausnutzung der Symmetrie führt zu der Vereinfachung, daß nur noch eine in einer Elementarzelle unbekannte Funktion u_k zu berechnen ist. Mit der Wellenzahl $\boldsymbol{k}$ kann man die Zustände klassifizieren. Zu verschiedenen $\boldsymbol{k}$ gehört ein verschiedenes Verhalten der Wellenfunktion bei einer Translation. Der Phasenfaktor ist jedoch gleich, wenn sich zwei Wellenzahlen $\boldsymbol{k}$ und $\boldsymbol{k}' = \boldsymbol{k} + \boldsymbol{G}$ um einen Vektor des durch $\boldsymbol{G}\cdot\boldsymbol{R} = 2\pi n$ (n = ganzzahlig) definierten *reziproken Gitters* unterscheiden. Die Gesamtheit aller Punkte des reziproken Gitters $\boldsymbol{G} = m_1\boldsymbol{b}_1 + m_2\boldsymbol{b}_2 + m_3\boldsymbol{b}_3$ ergibt sich aus ganzzahligen Vielfachen dreier Basisvektoren $\boldsymbol{b}_1 = 2\pi\dfrac{\boldsymbol{a}_2\times\boldsymbol{a}_3}{\boldsymbol{a}_1\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)}$, $\boldsymbol{b}_2 = 2\pi\dfrac{\boldsymbol{a}_3\times\boldsymbol{a}_1}{\boldsymbol{a}_1\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)}$, $\boldsymbol{b}_3 = 2\pi\dfrac{\boldsymbol{a}_1\times\boldsymbol{a}_2}{\boldsymbol{a}_1\cdot(\boldsymbol{a}_2\times\boldsymbol{a}_3)}$, die durch die Basisvektoren im Ortsraum bestimmt sind und ihrerseits eine Elementarzelle im $\boldsymbol{k}$-Raum aufspannen. Als *Brillouin-Zone* (Abb. 6.3.1–1) bezeichnet man die zugehörige Elementarzelle, ihr Volumen ist $\Omega_{\mathrm{BZ}} = (2\pi)^3/\Omega_{\mathrm{EZ}}$. Ein krz-Kristall hat ein kfz-reziprokes Gitter und umgekehrt.

krz: $\boldsymbol{b}_1 = (0, 1, 1)\, 2\pi/a$, $\boldsymbol{b}_2 = (1, 0, 1)\, 2\pi/a$,
$\boldsymbol{b}_3 = (1, 1, 0), 2\pi/a$,
kfz: $\boldsymbol{b}_1 = (-1, 1, 1)\, 2\pi/a$, $\boldsymbol{b}_2 = (1, -1, 1)\, 2\pi/a$,
$\boldsymbol{b}_3 = (1, 1, -1)\, 2\pi/a$,
hdp: $\boldsymbol{b}_1 = (2/\sqrt{3}, 0, 0)\, 2\pi/a$, $\boldsymbol{b}_2 = (1/\sqrt{3}, 1, 0)\, 2\pi/a$,
$\boldsymbol{b}_3 = (0, 0, 1)\, 2\pi/c$.

In der Kristallografie arbeitet man gern mit den Basisvektoren $\boldsymbol{a}_i^* = \boldsymbol{b}_i/2\pi$ für das reziproke Gitter. Für diese gilt $\boldsymbol{a}_i^*\cdot\boldsymbol{a}_j = \delta_{ij}$ statt $\boldsymbol{b}_i\cdot\boldsymbol{a}_j = 2\pi\delta_{ij}$.

6.3.2. Punktsymmetrie

Punkttransformationen sind *Spiegelungen*, *Drehungen* um eine (zwei-, drei, vier- oder sechszählige) Achse und die *Inversion*. Bei ihnen bleibt im Unterschied zu den Translationen mindestens ein Raumpunkt invariant. Aus der Punktsymmetrie eines Kristallgitters lassen sich gruppentheoretisch Aussagen über die möglichen Symmetrietypen (Darstellungen) von Elektronenwellenfunktionen und über die Entartung von Energiebändern machen. Aussagen dieser Art sind von einfachen Fällen her schon bekannt. Für

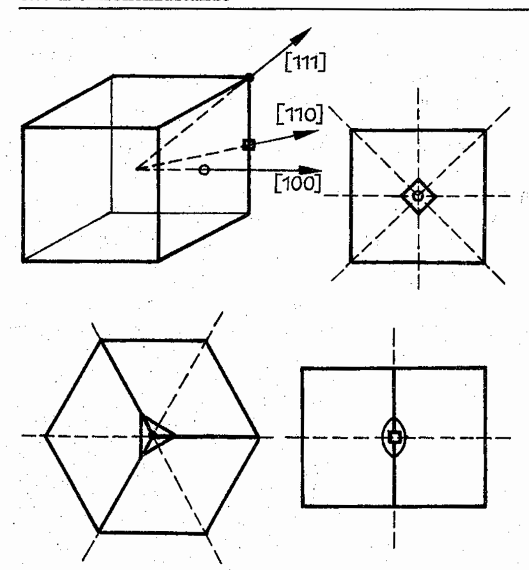


Abb. 6.3.2–1 Bei der Sicht in Richtung [100], [110] bzw. [111] erkennbare Symmetrieelemente des Würfels

$V(-x) = V(x)$ gibt es nur symmetrische und antisymmetrische Eigenfunktionen. In einem kugelsymmetrischen Potential ist ein Zustand mit dem Drehimpuls l immer $(2l + 1)$-fach entartet.
Die Zahl der voneinander verschiedenen Punktsymmetrien P eines Kristallgitters gibt an, wieviel physikalisch äquivalente Gebiete es in der Elementarzelle gibt. Das gleiche gilt für die Brillouin-Zone. Transformiert P einen Wellenvektor $\boldsymbol{k}$ in $\boldsymbol{k}' = P\boldsymbol{k}$, dann gilt für die Eigenwerte $E(\boldsymbol{k}) = E(\boldsymbol{k}')$.
Ein *kubischer Kristall* (krz, kfz, einfach kubisch) hat 48 Symmetrieelemente (Abb. 6.3.2–1). Die Energiebänder brauchen nur in dem in Abb. 6.3.1–1 hervorgehobenen 1/48 der Brillouin-Zone berechnet zu werden. Für den Wellenvektor $\boldsymbol{k} = 0$ (den Punkt Γ der Brillouin-Zone) gibt es zehn verschiedene Symmetrietypen von Wellenfunktionen, darunter zwei- und dreidimensionale Darstellungen. Im Falle $\boldsymbol{k} \neq 0$ wird durch $\boldsymbol{k}$ eine Richtung ausgezeichnet. Dadurch vermindert sich die für die Berechnung der Wellenfunktion verwertbare Punktsymmetrie. Die für $\boldsymbol{k} = 0$ bestehende Entartung wird (teilweise) aufgehoben. Bringt man ein Atom, dessen Elektronenzustände Lösungen in einem kugelsymmetrischen Potential – und damit $(2l + 1)$-fach entartet – sind, in ein Feld mit *kubischer Symmetrie*, dann wird diese Entartung teilweise aufgehoben. Mit der bei kubischer Symmetrie üblichen Bezeichnung der Darstellungen ist ein s-Zustand ($l = 0$) ein Γ_1-Zustand. Die drei p-Zustände ($l = 1$) haben die Symmetrie der dreidimensionalen Γ_{15}-Darstellung. Die fünf d-Zustände ($l = 2$) spalten in die beiden Darstellungen $\Gamma_{25'}$ (dreidimensional) und Γ_{12} (zweidimensional) auf.

6.3.3. Energiebänder bei starker Bindung

Man kann sich vorstellen, daß isolierte Atome bis zu dem experimentell beobachteten Gitterabstand a_e zusammengerückt werden. Die diskreten Niveaus der Atome gehen dabei in die Energiebänder des Festkörpers über. Abb. 6.3.3–1 zeigt einen Potentialtopf mit zwei gebundenen Zuständen und die zugehörigen Wellenfunktionen als eindimensionales Modell eines isolierten Atoms. Bei großem Abstand solcher Atome ist die Wechselwirkung gering, die Wellenfunktionen werden nur wenig deformiert. Abb. 6.3.3–2 verdeutlicht Wellenfunktionen mit zu $k = 0$ und $k = \pi/a$ gehörenden Translationseigenschaften ($e^{ika} = 1$ bzw. $e^{ika} = -1$). Die beiden aus φ_0 hervorgegangenen Zustände gehören zu den Energien $E_0 - \delta E$ und $E_0 + \delta E'$, für die die Wellenfunktion (vgl. Abb. 6.3.3–1) bei $a_e/2$ eine waage-

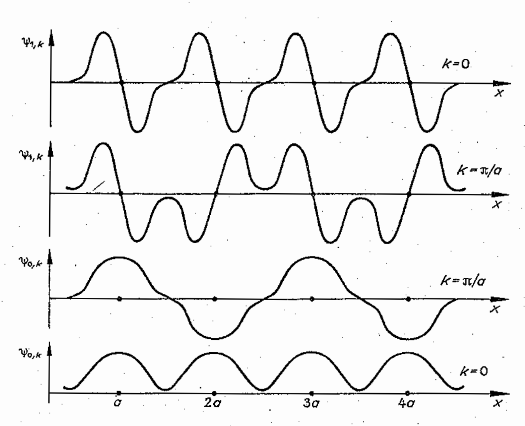


Abb. 6.3.3–2 Bloch-Funktionen im Bild starker Bindung

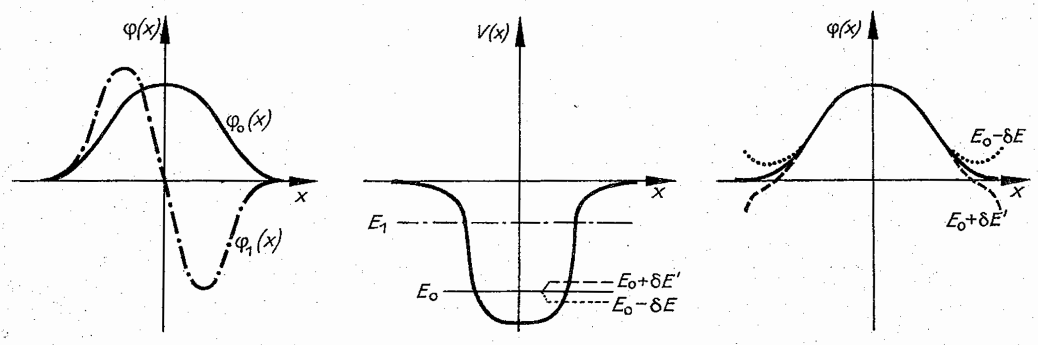


Abb. 6.3.3–1 Wellenfunktionen im Potentialtopf

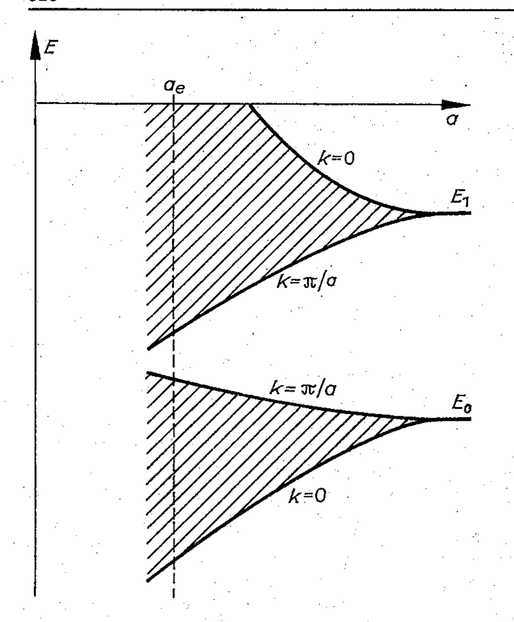


Abb. 6.3.3–3 Energiebänder in Abhängigkeit vom Gitterparameter

rechte Tangente bzw. einen Knoten hat. Die Energien der Zustände mit $0 < k < \pi/a$ liegen zwischen diesen beiden Werten, sie bilden ein *Energieband*, das bei Verringerung von a breiter wird (Abb. 6.3.3–3). Analoges gilt für die Energien der aus φ_1 hervorgehenden Zustände. Abb. 6.3.3–2 macht auch die *Energielücke* zwischen den Bändern verständlich. Während sich die Wellenfunktion von Zustand zu Zustand innerhalb eines Bandes wenig und kontinuierlich ändert, tritt beim Übergang zum nächsten Band eine sprunghafte Änderung auf.

Im Dreidimensionalen ist analog zu verstehen, wie sich die Energiebänder herausbilden. Nur treten hier eine größere Mannigfaltigkeit von Möglichkeiten und kompliziertere Verhältnisse durch die Entartung der atomaren Niveaus auf. Die zu verschiedenen atomaren Niveaus gehörenden Bänder können sich überlappen. Die beiden (Spin auf und Spin ab) Elektronen, die in gleichen diskreten Niveaus in den isolierten Atomen sitzen, haben auch in dem daraus entstehenden Energieband Platz, bei N Elementarzellen des Festkörpers (mit je einem Atom) also $2N$ Elektronen. Die Zahl der $\boldsymbol{k}$-Zustände in einem Band ist demzufolge gleich der Anzahl der Elementarzellen N des Festkörpers. Die Vereinbarung, daß man sich den Festkörper aus einer endlichen Anzahl N von Elementarzellen aufgebaut denkt und periodisch Randbedingungen einführt, hat zur Folge, daß es N Zustände gibt, die sich in ihrem Translationsverhalten unterscheiden. Die zugehörigen Werte der Quantenzahl $\boldsymbol{k}$ bilden ein *quasikontinuierliches Netz von Stützpunkten* in der Brillouin-Zone (N ist ja von der Größenordnung 10^{23}). Eine Summe über diese Zustände kann daher auch als Integral geschrieben werden. Die Umrechnung erfolgt über $\frac{1}{N}\sum_{k}^{BZ} = \frac{1}{\Omega_{BZ}}\int_{BZ} d^3k$. Unter *Gittersummenregeln* versteht man Relationen wie $\frac{1}{N}\sum_{R} e^{ik\cdot R} = \sum_{G} \delta_{k,G}$, $\frac{1}{N}\sum_{k}^{BZ} e^{ik\cdot R} = \delta_{R,0}$.
In dem in Abb. 6.3.3–2 dargestellten Fall starker Bindung kann eine Bloch-Funktion $\psi_{n,k}$ aus den an einem Atom lokalisierten, einer atomaren Wellenfunktion ähnlichen Zuständen φ_n aufgebaut werden,

$$\Psi_{n,k}(\boldsymbol{r}) = \sum_{R} e^{ik\cdot R}\varphi_n(\boldsymbol{r}-\boldsymbol{R}),\quad \Psi_k(\boldsymbol{r}) = \sum_{n} c_n(\boldsymbol{k})\,\Psi_{n,k}(\boldsymbol{r}).$$

Die Lösung der Schrödinger-Gleichung Ψ_k ist eine Linearkombination solcher Bloch-Funktionen. Die Eigenwerte eines Bandes haben die Gestalt $E_k = \sum_{R} e^{ik\cdot R} E_R$. Die Koeffizienten E_R enthalten Überlappungsintegrale über Wellenfunktionen und Kristallpotential an verschiedenen Gitterplätzen, mit wachsendem R werden sie rasch kleiner.

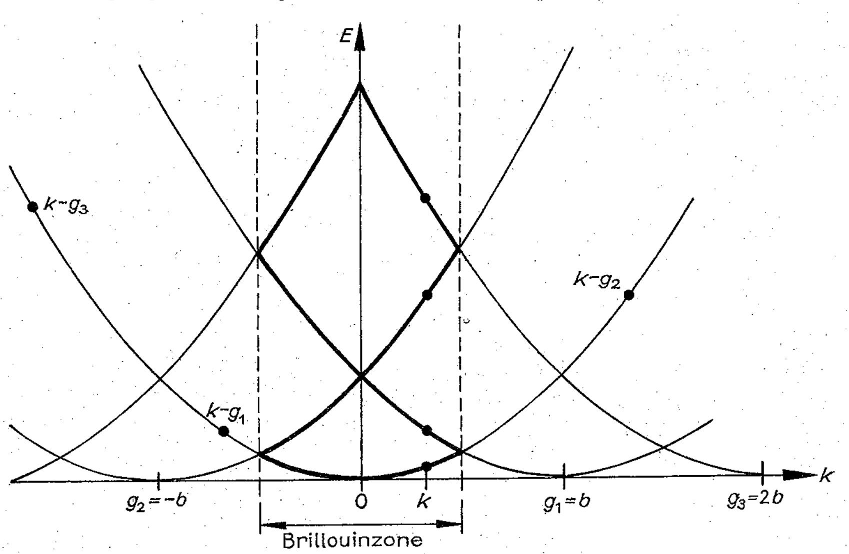


Abb. 6.3.4–1 Bandstruktur freier Elektronen, g_1, g_2, g_3 = Punkte des reziproken Gitters

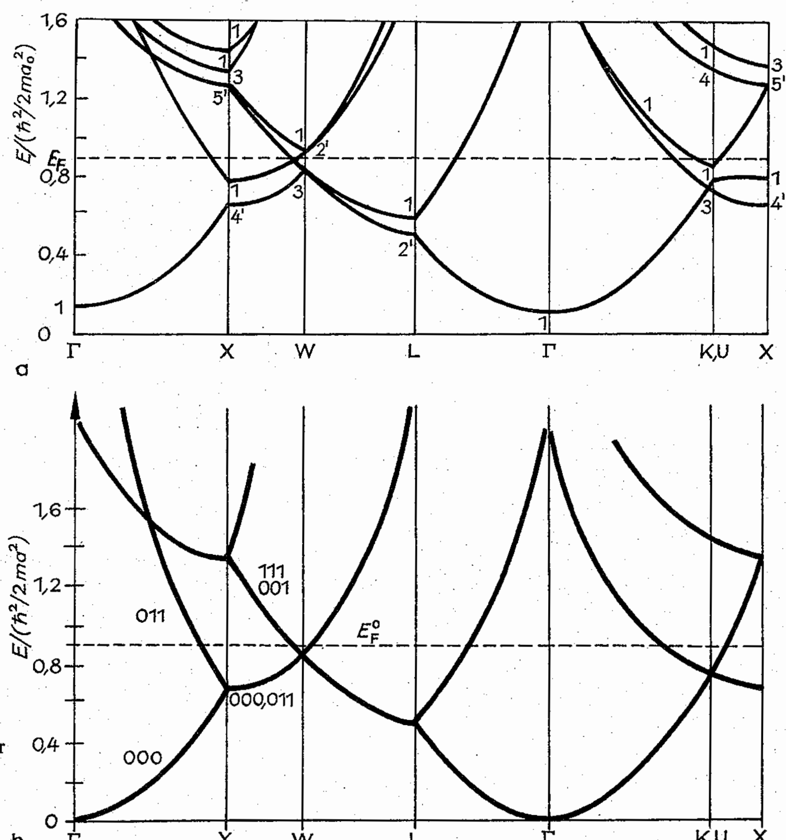


Abb. 6.3.4–2 (a) Bandstruktur von Aluminium, (b) im Bild freier Elektronen, 011 = Parabel um $G = b_2 + b_3$, usw. ($\hbar^2/2ma_0^2 = 13{,}6$ eV)

6.3.4. Energiebänder fast freier Elektronen

Man kann sich vorstellen, daß die atomaren Potentialtöpfe immer flacher werden, so daß das Potential schließlich in einen konstanten Wert übergeht. Welche Form haben in diesem Grenzfall die Energiebänder? Es zeigt sich, daß für Metalle, deren Leitungselektronen im isolierten Atom s- und p-Elektronen sind, dieser Grenzfall freier Elektronen, als *Bandstruktur freier Elektronen* bezeichnet, ein guter Ausgangspunkt für die tatsächliche Bandstruktur ist.

6.3.4.1. Freie Teilchen. Für sie ist der Impuls p eine Erhaltungsgröße, die Wellenfunktionen sind ebene Wellen $\varphi_p = 1/\sqrt{\Omega} \cdot e^{ip \cdot r/\hbar}$, die Energie ist gleich der kinetischen Energie $p^2/2m$. Um diese Zustände im Bild der Bandstruktur zu beschreiben, wird $p/\hbar = k - G$ in einen Vektor des reziproken Gitters und einen in der Brillouin-Zone liegenden Rest unterteilt. Die ebene Welle ist also bezüglich Translationen um Gittervektoren R eine zur Wellenzahl k gehörende Bloch-Funktion Ψ_k. Abb. 6.3.4–1 zeigt, wie die Parabel freier Elektronen für beliebige Wellenzahlen $p/\hbar$ auf die Brillouin-Zone reduziert wird. Man muß um jeden Punkt des reziproken Gitters G die Pararabel $(\hbar^2/2m)$ $(k - G)^2$ einzeichnen. Aus den in der Brillouin-Zone liegenden Teilen bauen sich die Energiebänder auf.

Im Dreidimensionalen wächst die Mannigfaltigkeit der Parabeln. Abb. 6.3.4–2 verdeutlicht, aus welchen Parabelteilen sich die Bandstruktur freier Elektronen für ein kfz-Gitter auf den in Abb. 6.3.1–1 erklärten Symmetrielinien zusammensetzt. Dem ist die tatsächliche Bandstruktur von Aluminium gegenübergestellt.

6.3.4.2. Fast freie Elektronen. Bei ihnen führen schwache atomare Potentialtöpfe an Punkten mit Entartung, d. h. Schnittpunkten von Parabeln, insbesondere an den Brillouin-Zonengrenzen, zu den wesentlichsten Abweichungen von dem Bild freier Elektronen. An den Brillouin-Zonengrenzen unterscheiden sich die Wellenzahlen der ebenen Wellen gleicher Energie um einen Vektor des reziproken Gitters. Diese ebenen Wellen erfüllen also die Laue-Bedingung, woraus auch zu ersehen ist, daß sie durch das Kristallgitter besonders stark beeinflußt werden. Der Bloch-Zustand wird dann eine Mischung aus ebenen Wellen, deren Wellenvektoren sich um Vektoren des reziproken Gitters unterscheiden, $\Psi_k(r) = \sum_G C_G(k)\, e^{i(k-G)r}$, wobei aber bei schwachen Potentialtöpfen nur wenige Summanden wesentliche Beiträge liefern. Die Schrödinger-Gleichung reduziert sich auf ein homogenes, lineares Gleichungssystem für die Koeffizienten C_G,

$$\sum_{G'} \left[\frac{\hbar^2 (k-G)^2}{2m} \delta_{G,G'} + V(G-G') \right] C_{G'}(k) = E(k) C_G(k),$$

in dem $V(G)$ *Fourier-Komponenten des Kristallpotentials* bedeuten. Die Eigenwerte, die Energiebänder $E(k)$ der Elektronen, ergeben sich aus der Forderung, daß die Koeffizientendeterminante dieses Gleichungssystems gleich Null ist.

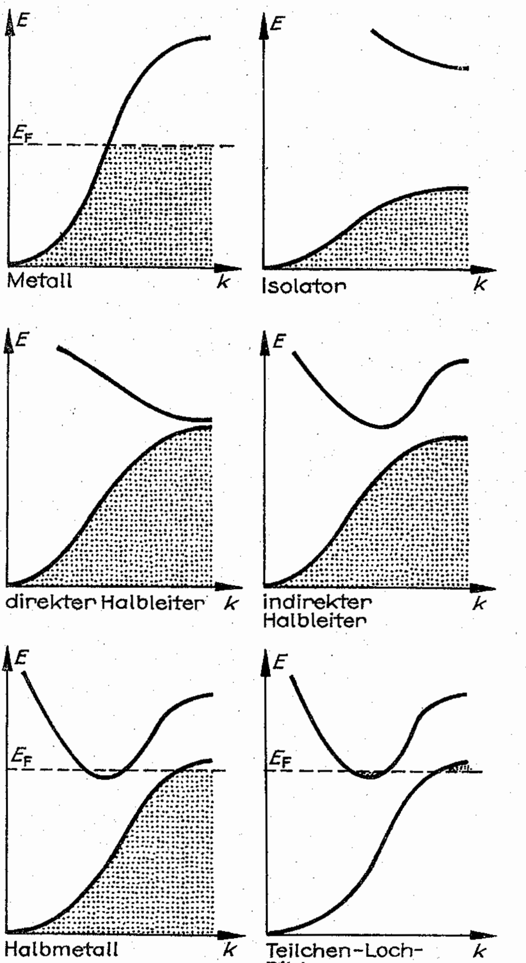


Abb. 6.3.5–1 Bandauffüllung und Festkörpertyp

Die Aufspaltung am Schnittpunkt zweier zu $\boldsymbol{G}$ bzw. $\boldsymbol{G}'$ gehörenden Parabeln ist bei einfachster Auswertung $\Delta E = 2\,|V(\boldsymbol{G}-\boldsymbol{G}')|$. Abb. 6.3.4–2 zeigt, wie das Kristallpotential die Parabelstücke freier Elektronen deformiert. Die Energiebänder sind periodische Funktionen $E(\boldsymbol{k}) = E(\boldsymbol{k}+\boldsymbol{G})$. Man kann sie daher über die erste Brillouin-Zone hinaus periodisch fortsetzen. Aus dem *reduzierten Zonenschema* wird dann das *wiederholte Zonenschema*.

Die Geschwindigkeit eines Wellenpaketes, die *Gruppengeschwindigkeit*, ist durch $v_{gr} = \partial E_k/\partial \hbar k$ gegeben. Für freie Elektronen $E = \hbar^2k^2/2m$ erhält man einfach $v_{gr} = \hbar k/m = p/m$. An den Brillouin-Zonengrenzen, wo die Energiebänder durch den Einfluß des Kristallpotentials waagerecht einmünden, wird die Gruppengeschwindigkeit Null, dort sind die Bloch-Zustände stehende Wellen.

6.3.5. Energiebänder von Festkörpern

Die Energiebänder $E_\nu(\boldsymbol{k})$ sind kontinuierliche Funktionen der *Wellenzahl* $\boldsymbol{k}$. Die Zustände eines Bandes werden durch $\boldsymbol{k}$-Werte innerhalb der ersten Brillouin-Zone unterschieden. Aus unterschiedlichen atomaren Niveaus gehen verschiedene Energiebänder hervor, die durch den *Bandindex* ν gekennzeichnet werden. Aus Abb. 6.3.5–1 geht hervor, wie die Überlappung der Bänder und die durch die Zahl der Valenzelektronen gegebene *Auffüllung der Bänder* den Typ des Festkörpers bestimmen. Das im Grundzustand höchste besetzte Niveau wird als *Fermi-Energie* E_F bezeichnet. Bei einem *Metall* gibt es unmittelbar über der Fermi-Energie unbesetzte Zustände, so daß Elektronen mit beliebig kleiner Energie angeregt werden können. Bei *Halbleitern* und *Isolatoren* muß dagegen eine *Energielücke* überwunden werden, um ein Elektron anzuregen. Beide unterscheiden sich nur durch die Größe der Energielücke. Sie liegt bei Halbleitern in der Größenordnung von Zehntelelektronenvolt, bei Isolatoren in der Größenordnung von Elektronenvolt. Bei Halbmetallen ist die Zahl der Zustände an der Fermi-Energie klein, ein Band ist fast gefüllt bzw. nur mit wenigen Elektronen besetzt.

6.3.5.1. Halbmetalle, Halbleiter. Bei Halbmetallen und bei Halbleitern, bei denen Elektronen aus dem Valenzbestand in das Leitungsband angeregt wurden, ist es einfacher, das *Teilchen-Loch-Bild* zu benutzen. Man arbeitet dabei nur mit

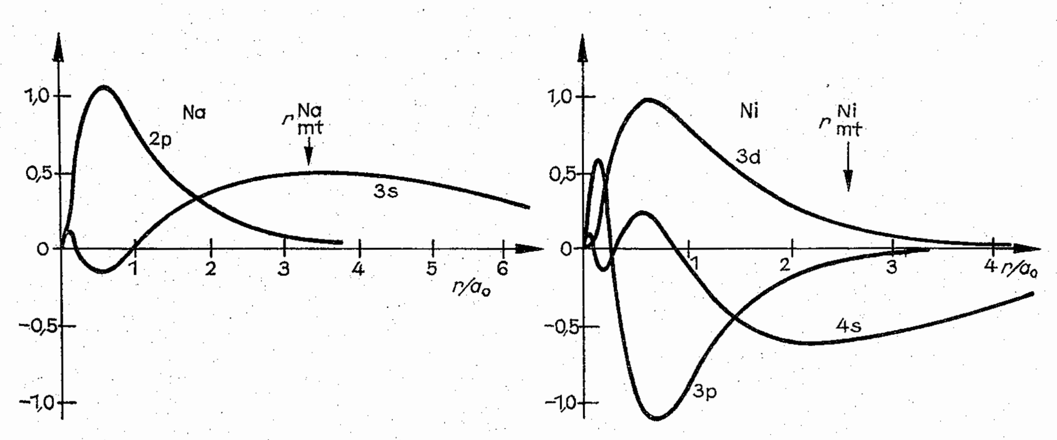


Abb. 6.3.5–2 Ausdehnung atomarer Wellenfunktionen im Vergleich zu r_{mt}

den (wenigen) unbesetzten Zuständen im Valenzband, anstatt die (vielen) Elektronenzustände zu betrachten. Für diese ist die Bandstruktur in der Umgebung von Minima und Maxima der Energiebänder (bei k_0) bedeutsam (vgl. 6.3.8.). Dort kann man den Bandverlauf durch eine Parabel annähern $E(k) = (\hbar^2/2m^*)(k - k_0)^2$, wobei man eine effektive Masse m^* einführt, um eine den freien Elektronen analoge Formel zu haben. m^* charakterisiert die Krümmung. Bei flachen Bändern ist die effektive Masse groß, bei steilen Parabeln klein. Da die Parabel in der Umgebung von Maxima nach unten gekrümmt ist, wird die effektive Masse negativ. Sie verliert ihre anschauliche Bedeutung. Geht man jedoch zur Teilchen-Loch-Beschreibung über, dann treten anstelle der unbesetzten Elektronenzustände besetzte Lochzustände mit positiver Ladung und positivem m^*.

6.3.5.2. Metalle. Bilden in ihnen die s- und p-Elektronen die Leitungselektronen, dann ist die Bandstruktur in dem Bild fast freier Elektronen (NFE-Bild) zu verstehen. Abb. 6.3.4–2 zeigt dies am Beispiel von Aluminium.

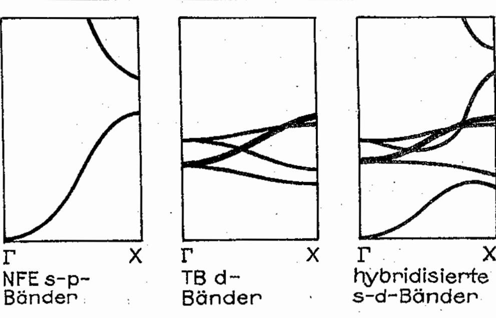


Abb. 6.3.5–3 H-NFE-TB-Bandstruktur

Bei *Übergangsmetallen* werden die Leitungsbänder durch die s- und d-Elektronen der obersten besetzten atomaren Niveaus gebildet. Die atomaren s-Wellenfunktionen sind im Vergleich zum halben Abstand r_{mt} nächster Nachbarn im

Abb. 6.3.5–4 Bandstruktur von (a) Vanadium und (b) Kupfer

Metall weit ausgedehnt (Abb. 6.3.5–2). Wellenfunktionen verschiedener Atome überlappen sich stark. Man erhält breite s-Bänder mit einer Bandstruktur, die im NFE-Bild verstanden werden kann. Die atomaren d-Wellenfunktionen sind dagegen relativ stark lokalisiert (vgl. Abb. 6.3.5–2), am Platz des nächsten Nachbarn sind sie weitgehend abgeklungen. Die zugehörigen Energiebänder können gut im Bild starker Bindung (TB-Bild) verstanden werden. Da sich die Bänder der s- bzw. d-Elektronen überlappen, tritt eine Mischung (*Hybridisierung*) beider Typen von Zuständen auf. Wie Abb. 6.3.5–3 veranschaulicht, ergibt sich die sogenannte *H-NFE-TB-Bandstruktur* [engl. *h*ybridized *n*early-*f*ree *e*lectron *t*ight *b*inding]. In Abb. 6.3.5–4 wird als konkretes Beispiel die Bandstruktur der Übergangsmetalle Vanadium und Kupfer dargestellt.

6.3.6. Kristallpotential

Unter dem Kristallpotential ist das Potential zu verstehen, das man in die Schrödinger-Gleichung für ein einzelnes Elektron einsetzen muß, um die Energiebänder und Bloch-Funktionen zu berechnen. Es wird aus dem Coulomb-Potential und dem Austausch- und Korrelationspotential gebildet.

6.3.6.1. Coulomb-Potential. Es ist aus der Ladungsdichte der positiven (Kerne) und negativen Ladungen (Elektronen) zu berechnen. Da die Ladungsdichte zunächst nicht bekannt ist, muß sie *selbstkonsistent* berechnet werden. Mit einer genäherten Ladungsdichte, die sich z.B. aus der Überlagerung atomarer Ladungsdichten ergibt, werden Bloch-Funktionen ausgerechnet, aus denen man eine verbesserte Ladungsdichte erhält. Aus dieser gewinnt man ein verbessertes Kristallpotential, bessere Bloch-Funktionen usw. Dieser Zyklus muß solange durchlaufen werden, bis sich der Bandverlauf oder das Potential innerhalb einer vorgegebenen Genauigkeit nicht mehr ändert.

6.3.6.2. Austausch- und Korrelationspotential V_{xc} sind quantenmechanische Effekte im Kristallpotential. Sie werden einerseits durch die *Antisymmetrie* der Gesamtwellenfunktion verursacht. Andererseits sind sie dadurch bedingt, daß die *paarweise Coulomb-Wechselwirkung* zwischen zwei Elektronen zunächst durch ein mittleres, nur die Ladungsdichte $n(r)$ enthaltendes Coulomb-Potential ersetzt wird. Das Austausch- und Korrelationspotential kann gut durch $V_{xc} = -\alpha(3/2)(3/\pi)^{1/3} e^2 \cdot n^{1/3}$ wiedergegeben werden, wobei der Parameter α zwischen 2/3 und 1 – je nach Substanz – liegt.

6.3.6.3. Streuphasen. Das Kristallpotential ist eine räumlich stark variierende Funktion. In der Umgebung jedes Atomkerns hat man einen tiefen, wie $-Ze^2/r$ verlaufenden Potentialtopf. Daraus erhebt sich die Frage, wieso das Bild fast freier Elektronen, d. h. von Elektronen, die sich in einem nahezu konstanten Potential bewegen, überhaupt richtige Energiebänder liefern kann. Um dies zu verstehen, muß man sich vor Augen halten, daß die Tiefe eines Potentials im wesentlichen die Anzahl der gebundenen Zustände bestimmt. Inwieweit ein einlaufendes freies Teilchen, eine ebene Welle, durch dieses Potential gestört wird, bestimmen aber die *Streuphasen* (vgl. 4.4.2.). Diese beschreiben den Verlauf der Wellenfunktion außerhalb des atomaren Potentialtopfes. Wenn – wie in Abb. 6.3.6–1a – die Streuphasen klein sind, d. h. die durch den atomaren Potentialtopf gestörte Wellenfunktion sich außerhalb des Potentialtopfes nicht stark von der Wellenfunktion eines freien Teilchens unterscheidet, dann hat auch das Kristallpotential nur einen geringen Einfluß auf die Bewegung eines Elektrons im Kristall. Das Elektron breitet sich nahezu wie ein freies Elektron aus. Berechnet man die Streuphasen für einen atomaren Potentialtopf in einem Al- oder Na-Kristall (vgl. Abb. 6.3.6–1a), dann findet man kleine Werte. Es wird verständlich, daß Aluminium und Natrium eine Bandstruktur ähnlich der freier Elektronen haben (vgl. 6.3.4–2). Die stärkste Streuung der Elektronen tritt bei einer Streuphase $\pi/2$ auf. Eine Streuphase π bedeutet nur einen Wechsel des unwesentlichen Vorzeichens der Wellenfunktion.

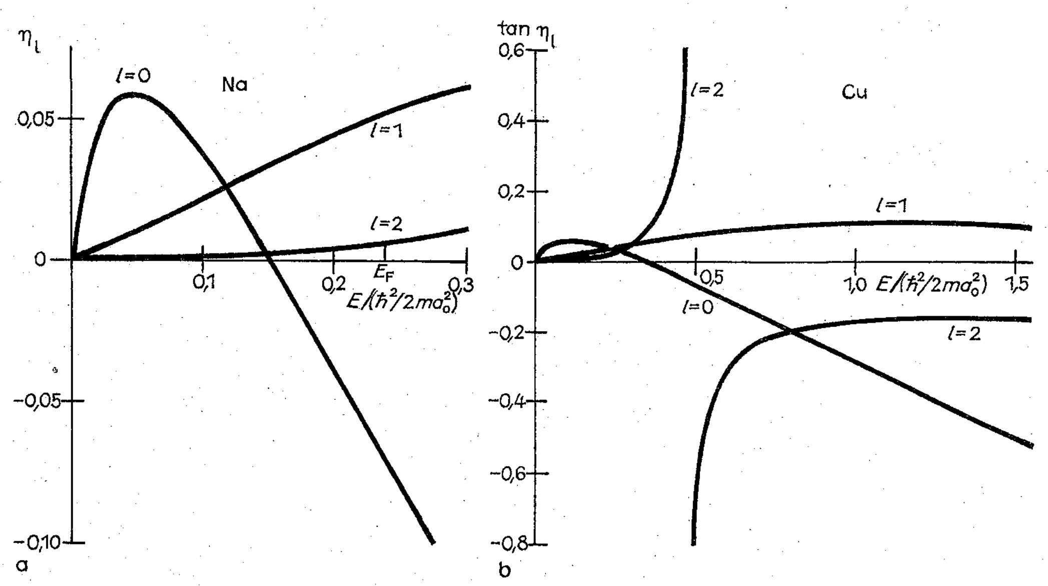


Abb. 6.3.6–1 Streuphasen η_l von (a) Natrium und (b) Kupfer

In *Übergangsmetallen* hat die d-Streuphase eine *Resonanz*, innerhalb eines kleinen Energiebereiches wächst sie von kleinen Werten auf π an, der Wert $\pi/2$ wird bei der Resonanzenergie durchlaufen (Abb. 6.3.6–1b). Aus dem Verlauf des Effektivpotentiales eines atomaren Potentialtopfes im Kristall wird dies verständlich. Der gebundene d-Zustand im isolierten Atom ist im Kristall ein virtuell gebundener Zustand.

6.3.6.4. Modellpotential. Die Bloch-Funktionen *einfacher Metalle* kann man nach *orthogonalisierten ebenen Wellen* (OPW-Funktionen) $\varphi^{\mathrm{OPW}}_{k+G} = \varphi_{k+G} - \varphi_c(\varphi_c, \varphi_{k+G})$ entwickeln. Dies sind im wesentlichen ebene Wellen $\varphi_{k+G}(r) = e^{i(k+G)\cdot r}$. Der Zusatzterm sorgt dafür, daß die OPW-Funktionen zu den atomaren Rumpfzuständen φ_c orthogonal sind. Die gesuchte Bloch-Funktion läßt sich als Linearkombination

$$\Psi_k = \sum_G c_G(k)\, \varphi^{\mathrm{OPW}}_{k+G} = \Psi^{\mathrm{NFE}}_k + \Psi^{\mathrm{Ortho}}_k$$

weniger OPW-Funktionen darstellen.

Man hat auch erfolgreich versucht, die Schrödinger-Gleichung so umzuformen, daß man eine Gleichung für Ψ^{NFE} selbst erhält. Zu dieser sogenannten *Pseudowellenfunktion* gehören die richtigen Eigenwerte; außerhalb des Rumpfbereiches stimmt Ψ^{NFE} auch mit der richtigen Bloch-Funktion überein (Abb. 6.3.6–2). Sie ist Lösung eines schwachen Kristallpotentiales, das aus flachen atomaren Potentialtöpfen besteht. Diese haben keine gebundenen Zustände und die im Rumpfbereich knotenfreie Funktion Ψ^{NFE} als Lösung. Hier kommt das Bild fast freier Elektronen voll zum Tragen. Das Pseudopotential ergibt sich aus dem Kristallpotential durch Abzug von Orthogonalisierungsanteilen. Oft benutzt man statt dessen *Modellpotentiale*, für die man einen glatten Verlauf ansetzt (vgl. Abb. 6.3.6–2). Die Parameter dieses Potentials können aus spektroskopischen Daten des isolierten Ions bestimmt werden.

Die Berechnung des Kristall- oder Modellpotentials ist der kritische Punkt bei jeder Bandberechnung. Bei First-principle-Rechnungen wird nur die Ordnungszahl vorgegeben. Wie erwähnt, kann man ein Modellpotential von atomaren Daten ausgehend gewinnen. Bei halbphänomenologischen Rechnungen paßt man einige Parameter an Meßdaten des Festkörpers an. Alle Wege sind erfolgreich beschritten worden.

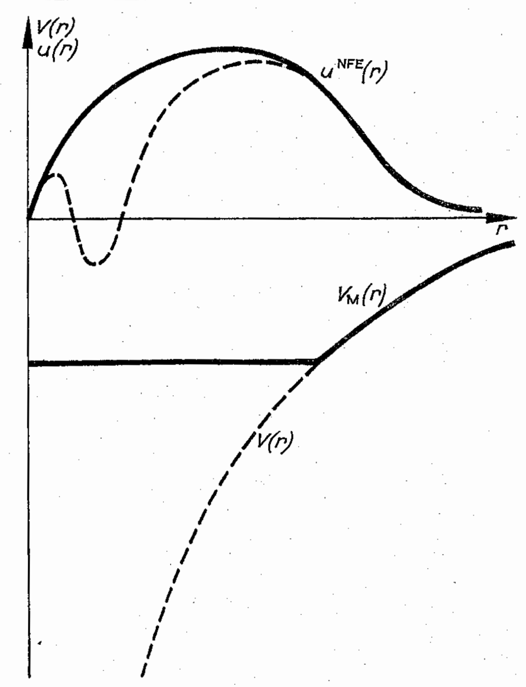


Abb. 6.3.6–2 Modellpotential $V_M(r)$ und Pseudowellenfunktion $u^{\mathrm{NFE}}(r)$ (———) im Vergleich mit Ionenpotential und -Wellenfunktion (– – –)

6.3.7. Aus der Bandstruktur abgeleitete Größen

Für die Diskussion vieler Meßgrößen ist es wichtig zu wissen, wieviele Zustände in einem Energieintervall $E \ldots E + \mathrm{d}E$ liegen. Für ein Energieband $E_\nu(k)$ sind das die Zustände zwischen den Flächen $E_\nu(k) = E$ und $E_\nu(k) = E + \mathrm{d}E$.

6.3.7.1. Zustandsdichte. Sie wird so normiert, daß die Anzahl der besetzten Zustände gleich der Zahl der Valenzelektronen je Elementarzelle Z_v ist,

$$D(E) = \frac{2}{\Omega_{\mathrm{BZ}}} \sum_\nu \int_{E_\nu(k)=E} \mathrm{d}^2k \frac{1}{|\partial E_\nu/\partial k|}, \quad N(E) = \int^E \mathrm{d}E\, D(E).$$

Die *integrierte Zustandsdichte* $N(E)$ hat an der Fermi-Energie E_F den Wert $N(E_F) = Z_v$ (vgl. 4.10.2.). Da der Nenner $\partial E/\partial k$ des Integrals verschwinden kann, weist die Zustandsdichte an Minima, Maxima oder Sattelpunkten der Bandstruktur sogenannte *Van-Hove-Singularitäten* (Abb. 6.3.7–1) auf.

Abb. 6.3.7–2 zeigt die Zustandsdichte einiger Substanzen. Bei Be wird man stark an die parabelförmige Zustandsdichte freier Elektronen $D(E) = (3/2) Z_v E_F^{-3/2} E^{1/2}$ erinnert. Für Übergangsmetalle ist die hohe Zustandsdichte im Bereich der flachen *d*-Bänder charakteristisch.

Zerlegt man in der Umgebung eines Atoms die Bloch-Funktionen in Anteile zu verschiedenen Drehimpulsen, dann erhält man die *partiellen Zustandsdichten*, die z. B. wesentlich die Röntgenemissionsspektren bestimmen. In einer Legierung definiert man *lokale Zustandsdichten*, die die Umgebung eines A- bzw. B-Atoms charakterisieren.

6.3.7.2. Fermi-Fläche. Sie ist die zur Fermi-Energie E_F gehörige Fläche konstanter Energie $E(k) = E_F$ im k-Raum. Die Elektronen an der Fermi-Fläche bestimmen das Verhalten des Festkörpers unter dem Einfluß äußerer Felder. Damit ein Elektron angeregt werden kann, muß es über die

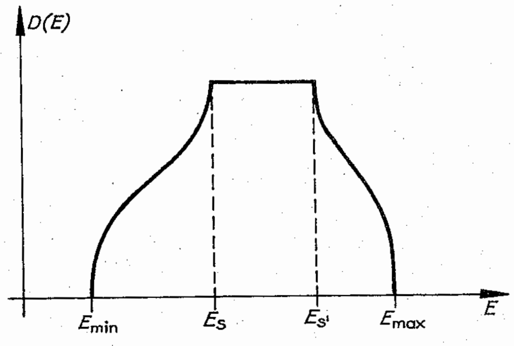


Abb. 6.3.7–1 Van-Hove-Singularitäten

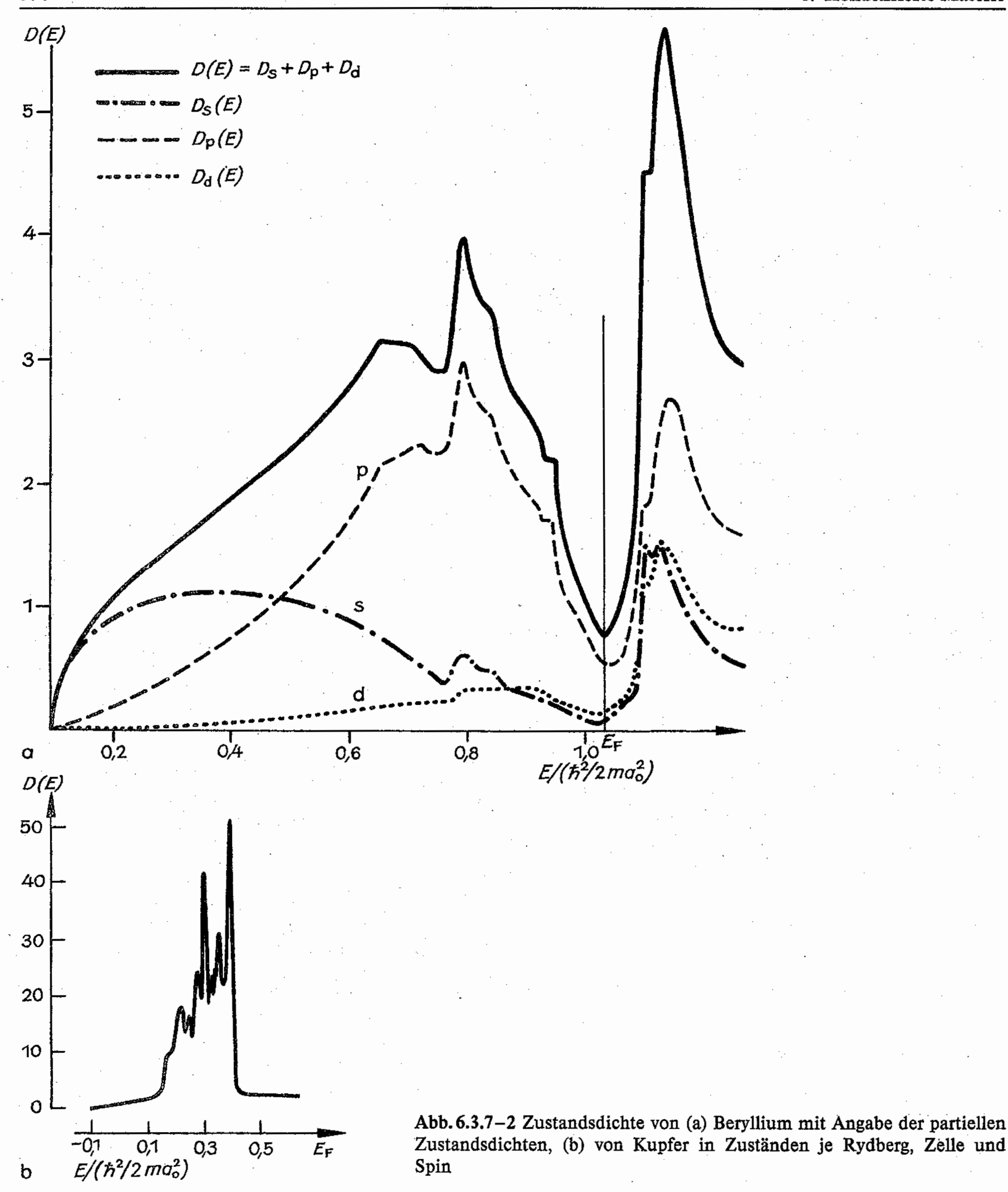


Abb. 6.3.7–2 Zustandsdichte von (a) Beryllium mit Angabe der partiellen Zustandsdichten, (b) von Kupfer in Zuständen je Rydberg, Zelle und Spin

Fermi-Energie gehoben werden, da es wegen des Pauli-Prinzips nur in einen unbesetzten Zustand übergehen kann. Durch kleine Anregungsenergien können daher nur Elektronen nahe der Fermi-Energie beeinflußt werden.

Abb. 6.3.7–3 zeigt die Fermi-Fläche von Kupfer. Sie ist im wiederholten Zonenschema ein im gesamten k-Raum zusammenhängendes Gebilde, man spricht von einer offenen Fermi-Fläche, sonst von einer geschlossenen. Umfaßt eine geschlossene Fermi-Fläche besetzte Zustände, dann spricht man von einer *Teilchenfläche*, umfaßt sie unbesetzte Zustände, von einer *Lochfläche*.

6.3.8. Energieniveaus in Halbleitern und Isolatoren

6.3.8.1. Besonderheiten der Bandstruktur. Grundsätzlich unterscheidet sich die Bandstruktur $E_\nu(\boldsymbol{k})$ von Halbleitern und Isolatoren nicht von der der Metalle (vgl. 6.3.5.); Besonderheiten sind jedoch

a) die *verbotene Zone* zwischen Valenz- und Leitungsband,

b) aus der geringen elektrischen Leitfähigkeit, die *langreichweitige elektrische Wechselwirkungen* ermöglicht, die in Metallen durch die frei beweglichen Ladungsträger abgeschirmt werden.

Die Darstellung der *Dispersionskurve* $E_\nu(k)$ längs ausgewählter, meist hochsymmetrischer Richtungen im k-Raum (Abb. 6.3.8–1) ist besonders zweckmäßig für die Diskussion optischer Prozesse (Absorption, Lumineszenz, Lichtstreuung), da daran meist mehrere Bänder beteiligt sind und die Quantenzahl k eine große Rolle spielt, die bei der Wechselwirkung von Elektronen mit Photonen und Phononen einem Erhaltungssatz genügt. Bei fast allen technisch wichtigen Halbleitern, Si, Ge, 3-5- bzw. 2-6-Verbindungen, liegt das Maximum des Valenzbandes bei $k = 0$. Dagegen liegt das (absolute) Minimum des Leitungsbandes bei sogenannten *direkten Halbleitern*, z. B. GaAs, InSb, ZnS, bei $k = 0$, bei den sogenannten *indirekten Halbleitern*, Si, Ge, GaP, bei $k \neq 0$.
Verzichtet man darauf, die Quantenzahl k in die grafische Darstellung einzubeziehen, so ist die *Zustandsdichte* $D_\nu(E)$ (vgl. 6.3.7.) des ν-ten Bandes die zur Veranschaulichung geeignete Größe. $D_\nu(E)\mathrm{d}E$ gibt die Zahl der erlaubten Elektronenzustände des ν-ten Bandes im Energieintervall zwischen E und $E + \mathrm{d}E$ an. $D_\nu(E)$ läßt sich aus der Bandstruktur $E_\nu(k)$ berechnen. Abb. 6.3.8–2 zeigt diesen Zusammenhang am Beispiel der vier Valenzbänder von GaAs.

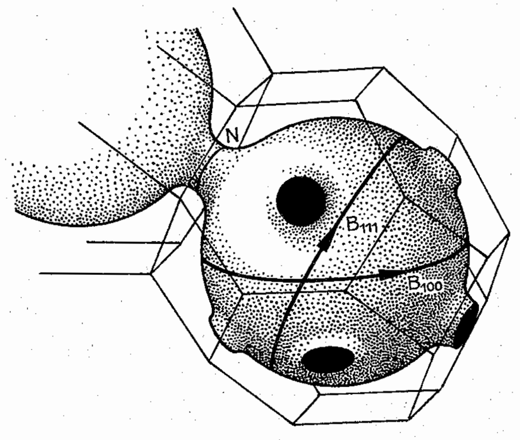


Abb. 6.3.7–3 Fermi-Fläche von Kupfer

Die Zustandsdichtedarstellung ist gut geeignet, Besetzungsfragen und Transportprobleme zu diskutieren, da sich die energetische Verteilung der *Elektronen* $n(E)$ bzw. der *Löcher* $p(E)$ (vgl. 6.3.5.) einfach in der Form $n(E) = D(E)f(E)$ bzw. $p(E) = D(E)[1 - f(E)]$ schreiben lassen. $f(E) = [1 + \exp(E - E_F)/kT]^{-1}$ ist die Fermi-Dirac-Verteilung (vgl. 4.10.7.). In dem interessierenden Temperaturbereich kann darin das chemische Potential durch die Fermi-Energie E_F ersetzt werden. Die Zustandsdichte behält ihre Bedeutung auch in ungeordneten Systemen, in denen das k-Raum-Konzept seinen Sinn verliert. Für optische Probleme ist sie dann angemessen, wenn die k-Auswahlregel keine Rolle spielt, z.B. bei Übergängen im Gebiet der Röntgen-Strahlen.
Bei Problemen des Ladungsträgertransportes in Halbleitern interessiert man sich meist nur für die Gesamtträgerkonzentrationen $n = \int n(E)\,\mathrm{d}E$ bzw. $p = \int p(E)\,\mathrm{d}E$. Für sogenannte *nichtentartete Halbleiter*, bei denen die Fermi-Energie E_F in der verbotenen Zone liegt und die Trägerverteilung in den Bändern demzufolge näherungsweise der Boltzmann-Statistik gehorcht, findet man $n = N_c e^{-(E_c - E_F)/kT}$ bzw. $p = N_v e^{-(E_F - E_v)/(kT)}$. Das bedeutet, daß sich Leitungs- bzw. Valenzband im nichtentarteten Halbleiter durch ein einheitliches Energieniveau E_c bzw. E_v ersetzen lassen, das dem unteren bzw. oberen Rand des Bandes entspricht. Die diesen Energieniveaus entsprechenden sogenannten *effektiven Zustandsdichten* $N_c = 2(2\pi m_c kT/h^2)^{3/2}$ bzw. $N_v = 2(2\pi m_v kT/h^2)^{3/2}$ sind temperaturabhängig; bei Zimmertemperatur

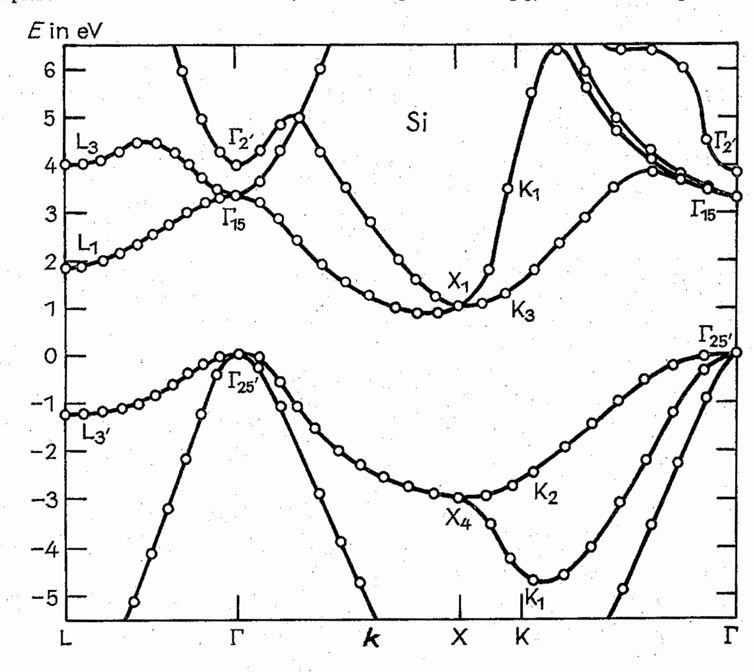


Abb. 6.3.8–1 $E(k)$-Kurve von Silicium längs ausgezeichneter Richtungen in der Brillouin-Zone Γ, L, X, K sind gruppentheoretische Symbole, die Punkte der Brillouin-Zone symbolisieren (vgl. Abb. 6.3.1–1b)

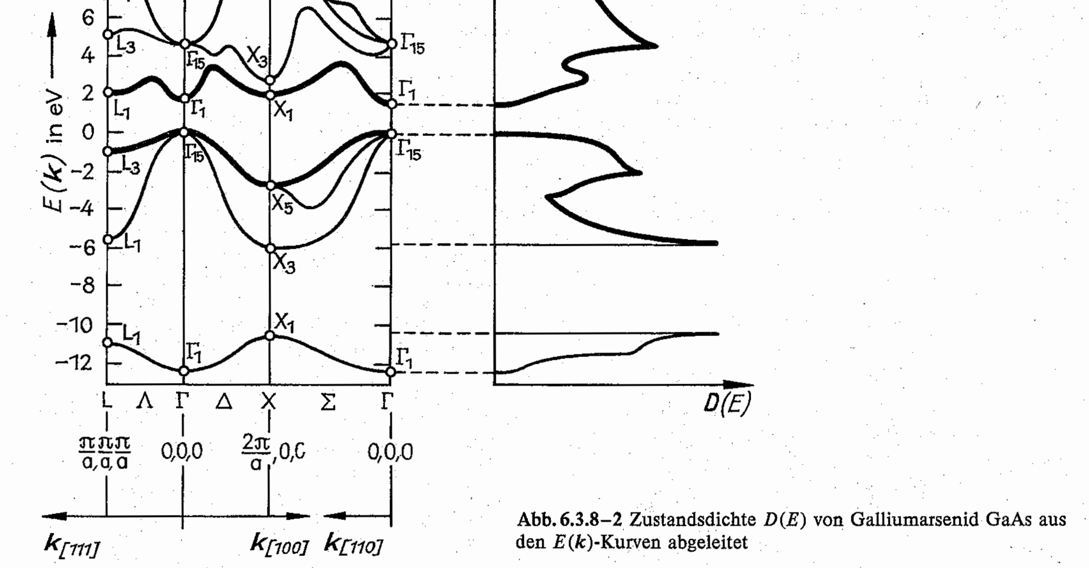


Abb. 6.3.8–2 Zustandsdichte $D(E)$ von Galliumarsenid GaAs aus den $E(\boldsymbol{k})$-Kurven abgeleitet

liegen sie in der Größenordnung von 10^{19} Zuständen/cm³ (h = Plancksches Wirkungsquantum). m_c und m_v sind die effektiven Massen der Träger in den betreffenden Bändern, die sich aus der Bandstruktur berechnen lassen (vgl. 6.3.5.) und in denen der Einfluß der Gitterbausteine auf die Bewegung der Elektronen bzw. Löcher enthalten ist. Sie sind i. allg. *anisotrop*, d. h. die Trägheit ist in unterschiedlichen Richtungen verschieden.

Diese Art der Veranschaulichung der Bandstruktur wird als *einfaches Bändermodell* bezeichnet und spielt vor allem bei der Erklärung der Wirkungsweise von Halbleiterbauelementen eine große Rolle, da man so in einfacher Weise örtliche Veränderungen der Bandstruktur darstellen kann.

Der Beschreibung der Elektronenbewegung durch die Bandstruktur liegt das *Einelektronenschema* zugrunde, d. h., jedes Elektron bewegt sich unabhängig von allen anderen in einem Potential, das sich aus dem der Atomrümpfe und dem gemittelten Potential aller anderen Elektronen zusammensetzt (vgl. 6.3.6.).

Berücksichtigt man die Elektron-Elektron-Wechselwirkung genauer, so wird man auf den Begriff des *Excitons* geführt: Regt man ein Elektron an, so daß es vom Valenz- ins Leitungsband übertritt, so kann wegen der Coulomb-Wechselwirkung zwischen Elektron und Loch im Valenzband ein gebundener Zustand entstehen, bei dem sich Elektron und Loch ähnlich wie Elektron und Proton im Wasserstoffatom um ihren gemeinsamen Schwerpunkt bewegen.

Die Energie eines derartigen Excitons beträgt in einfachen Fällen $E_n = E_g - \mu e^4/[2(4\pi\varepsilon\varepsilon_0)^2 \hbar^2 n^2] = E_g - R_{ex}/n^2$ (μ = reduzierte Masse ($1/\mu = 1/m_c + 1/m_v$), ε = Dielektrizitätskonstante des Kristalles, ε_0 = elektrische Feldkonstante, $\hbar = h/2\pi$), E_g = Breite der verbotenen Zone, $n = 1, 2, 3, \ldots$; die Hauptquantenzahl $n = 1$ beschreibt den Grundzustand des Excitons) (vgl. 4.2.2.).

Der Grenzfall $n \to \infty$ entspricht der Ionisation (Dissoziation) des Excitons in ein freies Elektron-Loch-Paar. Die Rydbergkonstante R_{ex} ist die Bindungsenergie im Grundzustand. Excitonen sind neutrale und bewegliche Anregungszustände eines Kristalles; bei einer Bewegung durch den Kristall wird Energie, aber keine Ladung transportiert. „Fällt" das Elektron in den unbesetzten Valenzbandzustand, das Loch, zurück, so spricht man von der *Rekombination* des Excitons. Dabei wird die Energie E_n frei, die als Lichtquant $h\nu$ abgestrahlt (strahlende Rekombination) oder in Wärme umgesetzt werden kann (strahlungslose Rekombination).

6.3.8.2. Experimentelle Methoden zur Bestimmung der

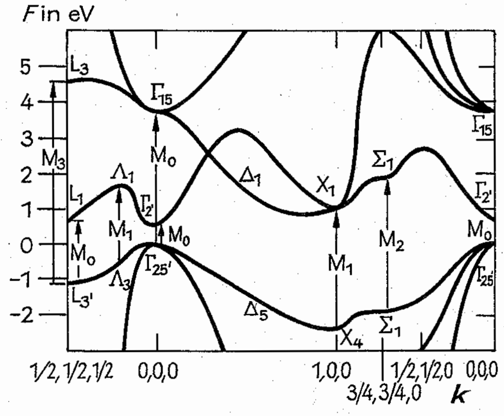


Abb. 6.3.8–4 Bandstruktur $E(\boldsymbol{k})$ von Germanium mit eingezeichneten kritischen Punkten M_0, M_1 und M_2

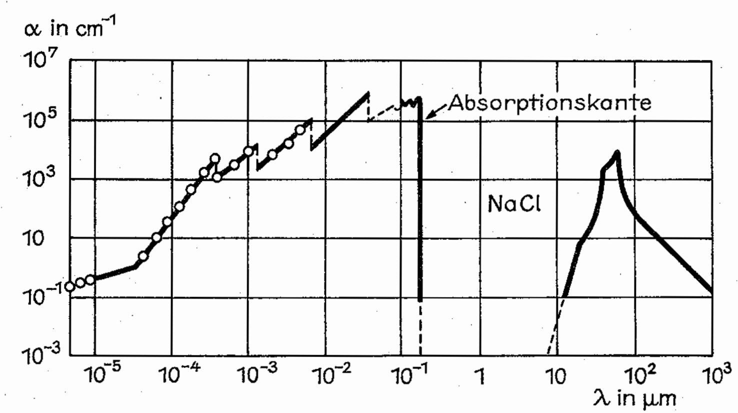


Abb. 6.3.8–3 Absorptionsspektrum von Natriumchlorid NaCl

Bandstruktur. Prinzipiell läßt sich die Bandstruktur eines Festkörpers bei vorgegebener chemischer Zusammensetzung und Gitterstruktur mittels quantenmechanischer Rechnungen bestimmen. Praktisch stehen dem aber beträchtliche Schwierigkeiten entgegen, so daß man meist zu halbempirischen Rechenverfahren greift, indem man einige typische Bandstrukturparameter dem Experiment entnimmt und die Rechnungen an diese Werte anpaßt. Für die Messung von Bandstrukturparametern stehen vorwiegend *optische Methoden* zur Verfügung.

Ein Absorptionsspektrum von NaCl – typisch für Nichtmetalle – zeigt Abb. 6.3.8–3. Der kurzwellige Teil (bis zur Absorptionskante) wird durch Elektronenübergänge verursacht und kann daher Aussagen zur Bandstruktur liefern; der langwellige Teil (im mittleren Infrarot) wird durch Gitterschwingungen verursacht.

Einen – wenn auch relativ groben – Gesamtüberblick über die Bandstruktur liefern die Absorptions- und Emissionsspektren im langwelligen Röntgen-Gebiet (Wellenlängen von ≈ 1 bis 30 nm, entsprechend Quantenenergien von 30 eV bis 1 keV). Absorption bzw. Emission von Lichtquanten dieser Energie entspricht Elektronenübergängen zwischen dem Leitungs- und Valenzband einerseits und tieferliegenden Elektronenzuständen andererseits. Letztere sind atomartig, d.h. nicht merklich zu Bändern verbreitert. Infolgedessen spielt die k-Auswahlregel keine Rolle und die Verhältnisse lassen sich in einem Zustandsdichteschema diskutieren.

Die *Absorption eines Röntgen-Quants* entspricht einem Übergang des Elektrons aus dem atomaren Energieniveau in einen leeren Endzustand, ist also erst möglich, wenn die Energie genügt, um das Leitungsband zu erreichen; der Absorptionskoeffizient ist daher proportional zur Zustandsdichte der (leeren) Leitungsbandzustände.

Die *Emission eines Röntgen-Quants* erfolgt beim Wiederauffüllen des bei der Absorption im atomaren Niveau entstandenen Loches, kann also nur von den (besetzten) Valenzbandübergängen aus erfolgen; die Emissionsintensität ist daher proportional zur Zustandsdichte im Valenzband. Die Differenz zwischen der niederenergetischen Kante des Absorptionsspektrums und der hochenergetischen Kante des Emissionsspektrums ist gleich der Breite der verbotenen Zone E_g.

Die *optische Absorption* im Spektralbereich von ≈ 0,1 µm bis zur Absorptionskante (bei Ge und Si im nahen Infrarot) wird durch Übergänge zwischen Valenz- und Leitungsband verursacht. Der Absorptionskoeffizient ist hier so hoch, daß meist das *Reflexionsspektrum* gemessen wird. Die Struktur des Spektrums wird vor allem durch die sogenannten *kritischen Punkte* (*Van-Hove-Singularitäten*) bestimmt; das sind die Stellen im k-Raum, an denen Valenz- und Leitungsband parallel zueinander laufen (Abb. 6.3.8–4). Eine genauere Identifizierung dieser kritischen Punkte ist mittels differentieller optischer Verfahren möglich. Beispielsweise mißt man bei der sogenannten *Elektroreflexion* die Differenz des Reflexionskoeffizienten mit und ohne von außen angelegtem elektrischem Feld und erhält charakteristische Strukturen an den kritischen Punkten (Abb. 6.3.8–5).

Detaillierte Aussagen über die Struktur der Bänder in der Umgebung der verbotenen Zone erhält man aus der Untersuchung der *Absorptionskante*. Eine sehr steile Kante bei hohem Absorptionskoeffizienten wird durch direkte Übergänge zwischen dem Valenzbandmaximum und einen im k-Raum am gleichen Ort befindlichen Leitungsbandminimum verursacht; eine flach verlaufende Kante bei kleinem Absorptionskoeffizienten entsteht durch Übergänge zwischen Bandextrema an unterschiedlichen Punkten des k-Raumes. Indirekte Übergänge erfolgen unter Beteiligung von Phononen (Gitterschwingungen; ihre genaue Analyse erlaubt daher auch die Messung von Gitterschwingungsspektren).

Bei einer Reihe von Isolatoren und Halbleitern tritt unter-

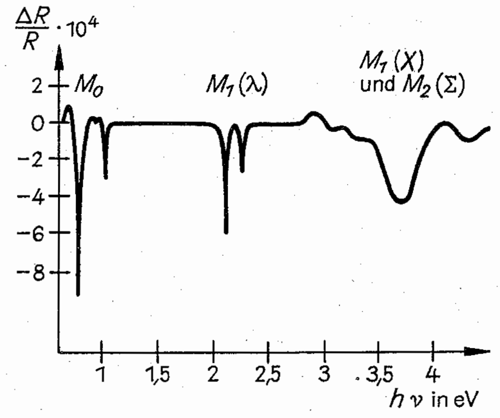


Abb. 6.3.8–5 Elektroreflexionsspektrum von Germanium, relative Änderung des Reflexionsvermögens R als Funktion der Photonenenergie

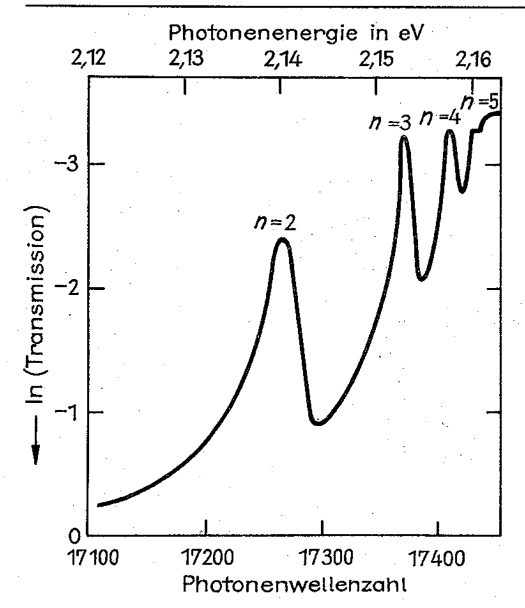


Abb. 6.3.8–6 Absorptionsspektrum von Kupferoxydul unterhalb der Bandbreite (fotografische Registrierung)

halb der Absorptionskante eine deutlich ausgeprägte linienhafte Absorption auf (Abb. 6.3.8–6); die Maxima entsprechen der Anregung der Excitonenniveaus E_n.

Sehr detaillierte Informationen über die Bandstruktur liefert die *Zyklotronresonanz* in einem äußeren Magnetfeld ***B***. Infolge der Lorentz-Kraft bewegen sich die Elektronen mit einer charakteristischen Umlauffrequenz (Zyklotronfrequenz) $\omega_c = |e| B/m_c$ auf Spiralbahnen um die Magnetfeldrichtung (m_c = effektive Elektronenmasse, $|e|$ = Betrag der Elektronenladung). Strahlt man eine elektromagnetische Welle ein, so wird diese im Falle der Resonanz stark absorbiert. Maxima der magnetfeldbedingten Absorption im Gebiet der Mikrowellen (oder im fernen Infrarot) ermöglichen so über die Messung von ω_c die Bestimmung der effektiven Masse, gleichbedeutend mit der Ermittlung der Krümmung im Minimum des Leitungsbandes bzw. – bei der Resonanz von Löchern – im Maximum des Valenzbandes.

Die Drehung des Kristalles im Magnetfeld verändert wegen der Anisotropie der effektiven Masse die Größe der Resonanzfrequenz und gestattet dadurch Aussagen über die Form der isoelektronischen Flächen.

6.3.8.3. Energieniveaus von Gitterstörungen in Halbleitern. Gitterstörungen können punktförmig (Leerstellen, Zwischengitter-, Fremdatome), linienförmig (Versetzungen) und flächenförmig (Oberflächen, Grenzflächen, Stapelfehler) sein. Sie führen stets zu einer Änderung der Zustandsdichte im Kristall, die sich vor allem bemerkbar macht, wenn *erlaubte Energieniveaus in der verbotenen Zone* der Bandstruktur entstehen. Änderungen der Zustandsdichte in den Bändern sind stets sehr klein, da die Zahl der Gitterstörungen viel kleiner als die der Wirtsatome ist.

Energieniveaus in der verbotenen Zone unterscheiden sich grundsätzlich von Bandzuständen; sie sind *lokalisierte Zustände*, d. h. die Aufenthaltswahrscheinlichkeit eines Elektrons in diesem Zustand ist auf die nähere Umgebung der Störung beschränkt.

Eine wichtige Folge dieser Lokalisierung ist bei Punktstörstellen, daß die Elektron-Elektron-Wechselwirkung eine sehr große Rolle spielt und daß durch die Bindung eines Elektrons an die Störstelle alle durch diese erzeugten Energieniveaus stark verändert werden, oft sogar vollständig verschwinden. Weitaus am wichtigsten für Halbleiter sind Fremdatome, die anstelle eines Wirtsatoms eingebaut sind und *Substitutionsstörstellen* heißen.

Ersetzt man z. B. ein Si-Wirtsatom (aus der vierten Gruppe des Periodensystems) durch ein Atom der fünften Gruppe, z. B. Phosphor, so werden vier der fünf Valenzelektronen des Störatoms für die chemische Bindung benötigt, d. h., sie besetzen Zustände im Valenzband. Es bleibt ein positiv geladener Atomrumpf zurück, in dessen elektrischem Feld sich das verbleibende fünfte Elektron bewegt. Die Situation ist der des Wasserstoffatomes völlig analog, so daß man wasserstoffähnliche, in der Umgebung der Störstelle lokalisierte Elektronenzustände erhält, deren Energien E_n durch (vgl. 4.2.2.) $E_n = E_c - E_B/n^2$ (n = 1, 2, 3, ...) gegeben sind. Die Hauptquantenzahl $n = 1$ entspricht dem Grundzustand, in dem das Elektron mit der Bindungsenergie $E_B = (1/\varepsilon^2)(m_c/m_0) R_H$ an den Atomrumpf gebunden ist (ε = Dielektrizitätskonstante, m_c = effektive Masse des Elektrons, m_0 = Masse des freien Elektrons, R_H = 13,6 eV = Rydberg-Konstante des Wasserstoffatomes).

Die Ausdehnung der Wellenfunktion im Grundzustand wird durch einen effektiven Bohrschen Radius $a_B = \varepsilon (m_0/m_c) a_0$ beschrieben (a_0 = 0,0529 nm ist der Bohrsche Radius des H-Atomes). In Halbleitern liegen die typischen Werte für E_B im Bereich von 10 bis 100 meV; die Bohrschen Radien sind mit etwa 1 bis 10 nm wesentlich größer als die Atomabstände im Gitter.

Bei genügend hoher Temperatur geht das Elektron aus dem gebundenen Zustand ins Leitungsband über (thermische Dissoziation, Ionisation der Störstelle) und wird frei beweglich. Diese Art von Störstellen wird als *Donator* (Elektronenspender) bezeichnet. Donatoren sind im leeren Zustand positiv geladen und in der Lage, Elektronen aus dem Leitungsband einzufangen.

Analoge Verhältnisse ergeben sich für Elemente der dritten Gruppe, z. B. Bor. Sie haben gegenüber dem Grundgitteratom ein Valenzelektron weniger, was einem „Loch" (Defektelektron) entspricht, das an die Störstelle gebunden ist. Bei höheren Temperaturen können die Störstellen Elektronen aus dem Valenzband des Wirtskristalles aufnehmen („Löcher ins Valenzband emittieren"); sie werden als *Akzeptoren* bezeichnet. Ein Akzeptor ist leer, wenn er ein Loch ins Valenzband abgegeben hat; das bedeutet, daß er ein zusätzliches Elektron aufgenommen hat und deshalb negativ geladen ist. Leere, d. h. negativ geladene, Akzeptoren können Löcher aus dem Valenzband einfangen.

Je nach dem dominierenden Störstellentyp spricht man von Halbleitern vom *n-Typ* (überwiegend Donatoren, negative Ladungsträger im Überschuß) oder *p-Typ* (überwiegend Akzeptoren, positive Ladungsträger im Überschuß). Sind Donatoren und Akzeptoren in vergleichbaren Konzentrationen vorhanden, so heißt der Halbleiter kompensiert (*k-Typ*). Der Eigen- oder Idealhalbleiter (ohne Störstellen, engl. „intrinsic") wird als *i*-Typ bezeichnet. Durch bewußtes Einbringen von Verunreinigungsatomen (Dotieren) während der Züchtung oder nachträglich durch Diffusion bzw. Ionenimplantation („Einschießen" von Ionen) läßt sich die Trägerkonzentration und damit die Leitfähigkeit in weiten Grenzen verändern, was die Grundlage für die Halbleitertechnologie ist.

Donatoren und Akzeptoren erzeugen meist sogenannte *flache Störstellenniveaus*, d. h., die Bindungsenergie des Elektrons bzw. Loches ist wesentlich kleiner als die Breite der verbotenen Zone. Donatoren stehen nur mit dem Leitungsband, Akzeptoren nur mit dem Valenzband in starker Wechselwirkung (vgl. Abb. 6.3.8–8).
Ähnlich wirken *Haftstellen*, auch sie stehen nur mit einem Band in starker Wechselwirkung. Beispielsweise kann eine im leeren Zustand neutrale Gitterstörung Elektronen einfangen, sie ist dann im besetzten Zustand negativ geladen und heißt Elektronenhaftstelle; analog ist eine Löcherhaftstelle im besetzten Zustand positiv geladen.
Außer diesen flachen gibt es *tiefe Störstellenniveaus*, die z. B. durch Leerstellen, Übergangsmetallionen u. a. verursacht werden. Sie stehen in starker Wechselwirkung sowohl mit dem Leitungs- als auch mit dem Valenzband und binden sowohl Elektronen als auch Löcher mit Bindungsenergien, die vergleichbar mit der Breite der verbotenen Zone sind. Energetisch liegen diese Niveaus etwa in der Mitte der verbotenen Zone. Sie können sowohl Elektronen als auch Löcher einfangen und sind daher sehr effektive *Rekombinationszentren*.
Außer einzelnen Ladungsträgern können Störstellen auch Excitonen einfangen. Derartige *gebundene Excitonen* sind dann unbeweglich. Sie rekombinieren oft unter Aussendung von Licht, wobei sehr scharfe Spektrallinien entstehen können (vgl. Abb. 6.3.8–11). Jede *Oberfläche* stellt eine sehr drastische Gitterstörung dar und führt zu sogenannten Oberflächenzuständen (vgl. 6.10.4.2.), deren Anzahl etwa gleich der Anzahl der Oberflächenatome ist. Ihr Vorhandensein kann zu beträchtlichen elektrischen Oberflächenladungen führen, die ihrerseits Raumladungsrandschichten in Halbleitern nach sich ziehen. Die Folge ist, daß die Oberfläche eines Halbleiters sich in elektrischer Hinsicht drastisch vom Volumen unterscheiden kann, wobei äußere Einflüsse, wie Gasbedeckung in Form von Adsorptionsschichten, Oberflächenoxide, eine große Rolle spielen. Die technologische Beherrschung der Oberflächeneigenschaften ist daher u. a. eine Grundvoraussetzung zur Herstellung von Halbleiterbauelementen. Ähnlich komplizierte und bisher nur teilweise geklärte Verhältnisse liegen an der Grenzfläche zweier verschiedener Festkörper, z. B. bei metallischen Kontakten für Halbleiter, vor. Die betreffenden lokalisierten Zustände heißen Grenzflächenzustände (*interface-states*).
Für die *meßtechnische Erfassung von Störstellen* stehen eine Vielzahl chemischer, elektrischer, optischer und hochfrequenz-spektroskopischer Verfahren zur Verfügung, deren Anwendung jedoch nicht universell ist, sondern meist auf bestimmte Störstellentypen oder Stoffgruppen beschränkt ist und nur bestimmte Parameter liefert. Die wichtigsten Störstellenparameter sind: die chemische Natur, die Konzentration, der Einbauort, die energetische Lage, der Umladungscharakter, die Einfangkoeffizienten für Elektronen und Löcher, die Emissionskoeffizienten bei thermischer Anregung, die Ionisationswahrscheinlichkeiten bei optischer Anregung.
Zur Ermittlung der *chemischen Natur* und der *Konzentration* der Störstellen können teilweise die Verfahren der analytischen Chemie oder der Massenspektroskopie herangezogen werden. Bei den für Halbleiter typischen geringen Konzentrationen reicht die Nachweisempfindlichkeit dieser Methoden jedoch oft nicht aus, so daß man auf mittelbare Informationen aus elektrischen Messungen angewiesen ist. In manchen Fällen gelingt es auch, durch Störstellen hervorgerufene Gitterschwingungen (*lokale Moden*) im Absorptionsspektrum nachzuweisen, aus deren Frequenz auf die Masse des schwingenden Störatoms geschlossen werden kann.
Konzentrationsbestimmungen von Donatoren und Akzeptoren am homogenen Material sind durch Messung der Leitfähigkeit und des Hall-Effektes (vgl. 8.1.1.1.) möglich, die primär die Elektronenkonzentration (im n-Halbleiter) bzw. die Löcherkonzentration (im p-Halbleiter) liefern. Bei genügend hohen Temperaturen, wenn alle Störstellen ionisiert sind, stimmt diese Konzentration n bzw. p der freien Ladungsträger mit der Störstellenkonzentration N_D bzw. N_A überein. Bei niedrigen Temperaturen ist im n-Halbleiter $n = N_D e^{-E_D/2kT}$, so daß man aus der Auftragung von ln n über $1/kT$ die Bindungsenergie E_D erhält, die mit der chemischen Natur des Donators zusammenhängt. Analoges gilt für p-Halbleiter.
Elektrische Messungen an tiefen Störstellen, insbesondere Haftstellen, sind mit der Methode thermisch stimulierter Ströme (engl. *t*hermally *s*timulated *c*urrent, *TSC*) möglich.

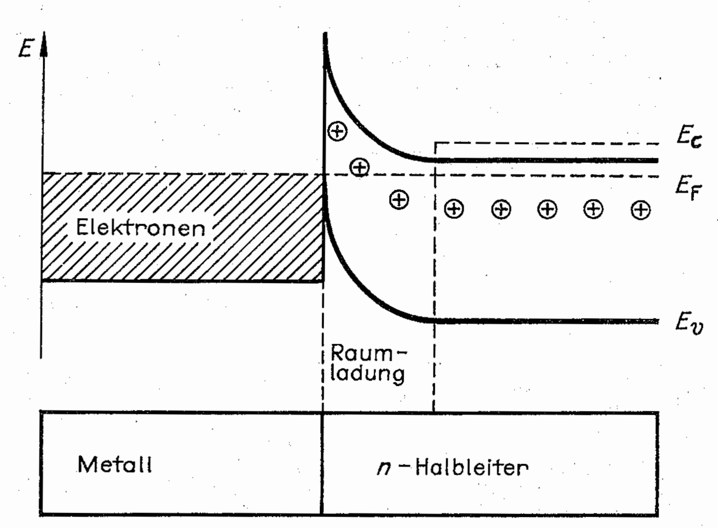


Abb. 6.3.8–7 Schema der Raumladungsrandschicht eines n-Halbleiters (Schottky-Barriere)

Dabei kühlt man das zu untersuchende Material auf möglichst tiefe Temperaturen ab und bestrahlt es mit Licht, das freie Elektron-Loch-Paare erzeugt, die in die Haftstellen (traps) eingefangen werden. Anschließend erhöht man allmählich die Temperatur, wodurch zunächst aus flachen, bei höheren Temperaturen auch aus tieferen Haftstellen Ladungsträger in die Bänder angeregt werden. Jede Haftstellensorte gibt Anlaß zu einem Strommaximum, dessen Höhe ihrer Konzentration proportional ist; die Temperatur, bei der das Maximum auftritt, erlaubt in günstigen Fällen die Bestimmung der Bindungsenergie der Ladungsträger an der betreffenden Störstelle. Ähnliche Schlußfolgerungen lassen sich aus der spektralen Abhängigkeit der Fotoleitfähigkeit ziehen.

Konzentrationsbestimmungen flacher Störstellen an *pn*-Übergängen oder Metall-Halbleiter-Kontakten (Schottky-Kontakten) sind mittels *Kapazitätsmessungen* bei Sperrpolung möglich. Als Beispiel betrachte man einen Schottky-Kontakt an einem *n*-Leiter (Abb. 6.3.8–7). Bei Sperrpolung (Metall negativ, Halbleiter positiv) entsteht im Halbleiter an der Kontaktfläche eine Raumladungsrandschicht positiv geladener Donatoren, da Elektronen ins Halbleiterinnere getrieben werden und wegen der Potentialbarriere aus dem Metall keine nachgeliefert werden. Die Raumladungsschicht stellt eine (spannungsabhängige) Kapazität C dar, die mit üblichen Kapazitätsmeßbrücken gemessen werden kann. Aus der Beziehung $C = A\sqrt{\varepsilon\varepsilon_0 e N_D/(2U_r)}$ (A = Kontaktfläche, ε = Dielektrizitätskonstante des Halbleiters, e = Elementarladung, U_r = Sperrspannung) läßt sich die Donatorkonzentration N_D berechnen. Ähnliche Überlegungen gelten auch für sperrgepolte *pn*-Übergänge.

Befinden sich in der Raumladungsschicht eines *pn*-Überganges oder eines Schottky-Kontaktes außer den flachen noch tiefe Störstellen, so führt deren Umladung (Füllung oder Entleerung) zu zeitlichen Änderungen der Sperrschichtkapazität. Die Umladung kann elektrisch, thermisch oder optisch erfolgen; mißt man dabei ähnlich wie bei der *TSC*-Methode die zeitlichen Änderungen der Kapazität als Funktion der Temperatur (die sogenannten Kapazitätstransienten), so erhält man daraus Informationen über Konzentration, energetische Lage und die Emissionskoeffizienten der tiefen Störstellen. Bei optischer Umladung durch Einstrahlung von Licht passender Wellenlänge kann man zusätzlich die optischen Ionisationswahrscheinlichkeiten ermitteln. Diese Verfahren der Untersuchung tiefer Störstellen werden unter dem Namen *DLTS* (engl. *d*eep *l*evel *t*ransient *s*pectroscopy) häufig angewendet. Sehr detaillierte Aussagen über die Energieniveaus von Störstellen erhält man aus *optischen Spektren*, die infolge von elektronischen Übergängen innerhalb einer Störstelle, zwischen Störstellen und Bändern und zwischen verschiedenen Störstellen erfolgen (Abb. 6.3.8–8). Diese Übergänge können teilweise im Absorptionsspektrum nachgewiesen werden, oft beachtet man sie aber auch im Emissionsspektrum als *Lumineszenzstrahlung*, die durch Licht (Fotolumineszenz), energiereiche Elektronen (Katodolumineszenz) oder elektrische Felder (Elektrolumineszenz) angeregt werden kann.

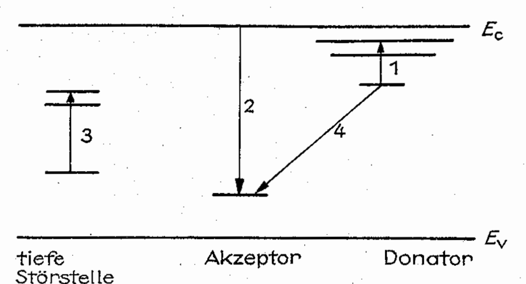


Abb. 6.3.8–8 Optische Übergänge unter Beteiligung von Störstellen; 1 Intra-Donator-Übergang, 2 Leitungsband-Akzeptor-Übergang, 3 Übergang innerhalb einer tiefen Störstelle, 4 Donator-Akzeptor-Übergang

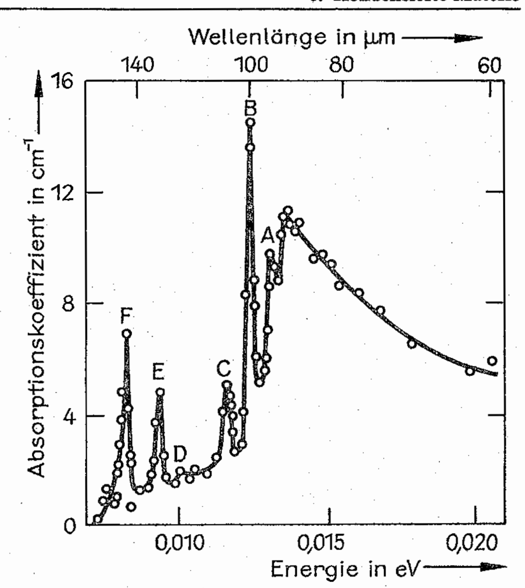


Abb. 6.3.8–9 Absorptionsspektrum von Arsendonatoren in Germanium. Die Linien entsprechen Übergängen vom Grundzustand in angeregte Zustände, das Kontinuum bei Energien oberhalb der mit A bezeichneten Linie Übergängen in Leitungsbandzustände

Als Beispiel für Übergänge innerhalb einer Störstelle zeigt Abb. 6.3.8–9 die linienhafte Absorption, die durch As-Störstellen in Germanium hervorgerufen wird. Das Spektrum liegt im IR-Spektralbereich und ist nur bei tiefen Temperaturen zu beobachten, wenn alle Elektronen am Donator gebunden sind. Die Lage der Linien bestätigt sehr gut das Modell des „verschmierten" Elektrons, auf dessen Grundlage die Energiezustände von Donatoren berechnet wurden. Das Spektrum steht in keinem physikalischen Zusammenhang mit dem des freien As-Atoms.

Im Gegensatz dazu ähnelt das Spektrum von Eu-Ionen in Y_2O_3 sehr stark dem des freien Eu^{3+}-Ions. Es wird durch Übergänge der Elektronen innerhalb der 4f-Schale von Eu^{3+}-Ionen verursacht, die auf Y^{3+}-Plätzen eingebaut sind. Die Energiezustände dieser relativ fest gebundenen Elektronen werden durch eine schwache Wechselwirkung mit den benachbarten O-Ionen aufgespalten (innerkristalliner Stark-Effekt). Die dadurch verursachten charakteristischen Feinstrukturen im Emissions- und Absorptionsspektrum gestatten es, durch Symmetriebetrachtungen u. a. die genaue Lage des Störatomes im Kristallgitter zu ermitteln. Da die Spektren solcher Ionen der Seltenen Erden bzw. der Übergangsmetalle aus sehr scharfen Linien bestehen, sind

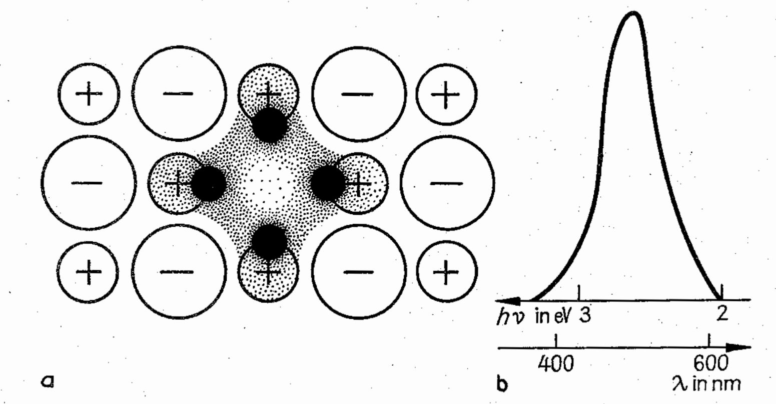


Abb. 6.3.8–10 (a) Aufenthaltswahrscheinlichkeit eines Elektrons an einer Anionenfehlstelle, (b) Absorptionsbande des F-Zentrums in NaCl

Stoffe mit derartigen Ionen gut als Materialien für Festkörperlaser geeignet; z. B. arbeitet der Rubinlaser mit mit Chromium dotiertem Al_2O_3.

In anderen Fällen ist die Wechselwirkung der Störstelle mit dem Wirtskristall weit stärker, so daß statt einzelner Linien einige breite Absorptions- oder Lumineszenzbanden entstehen. Typische Beispiele sind die Spektren der sogenannten *Farbzentren* (*F*-Zentren) in Alkalihalogeniden, die durch den Einfang angeregter Elektronen in Anionenfehlstellen entstehen (Abb. 6.3.8–10). Der entsprechende Elektronenübergang ist dabei stets von einer mehr oder weniger starken Änderung des Schwingungszustandes des Kristallgitters begleitet. Es werden gleichzeitig mit der Umwandlung von Licht- in elektronische Energie eine oder mehrere Phononen erzeugt oder vernichtet.

Als Beispiel für einen *Band-Störstellen-Übergang* zeigt Abb. 6.3.8–11 die durch Leitungsband-Akzeptor-Übergänge verursachte Lumineszenz von *p*-leitendem Silicium. Die entsprechenden Banden sind relativ breit, da die Elektronen eine mittlere thermische Energie von $3/2\,kT$ haben. Aus der spektralen Lage dieser Banden kann man näherungsweise den energetischen Abstand zwischen Leitungsband und Akzeptorniveau bestimmen.

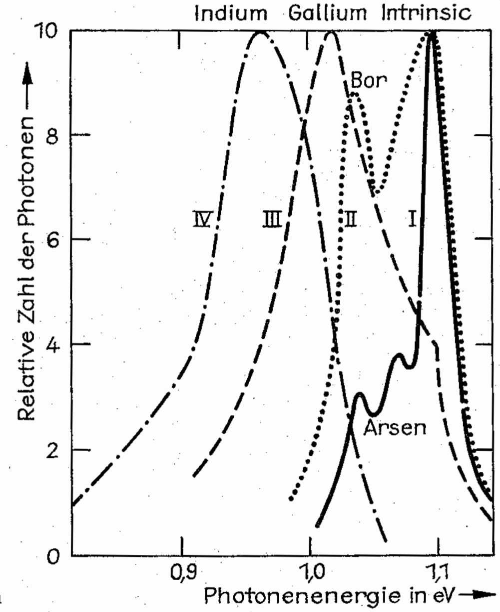


Abb. 6.3.8–11 Lumineszenzspektren von *p*-leitendem Silicium

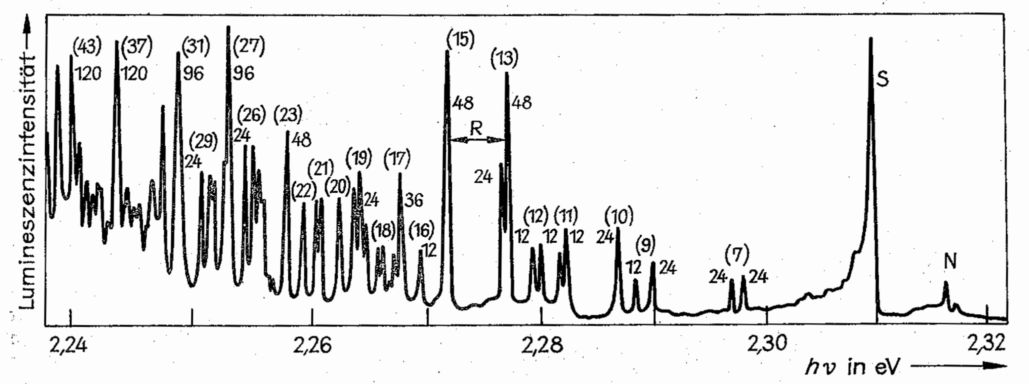


Abb. 6.3.8–12 Donator-Akzeptor-Paarspektrum von Schwefel und Kohlenstoff in Galliumphosphid bei 1,5 K. Die Zahlen in Klammern numerieren die Linien, die übrigen Zahlen sind die theoretisch zu erwartenden Intensitäten; S und N siehe Text.

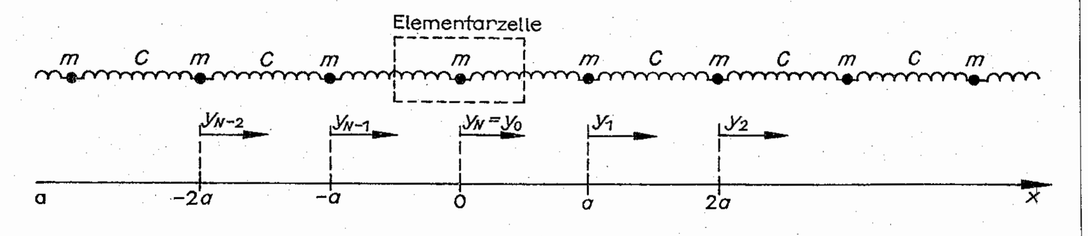
Zwischenstörstellenübergänge verursachen in einigen Halbleitern außerordentlich informationsreiche Spektren. Abb. 6.3.8–12 zeigt das sehr linienreiche Lumineszenzspektrum von GaP, das mit Schwefel (Donator) und Kohlenstoff (auf Phosphorplatz, deshalb Akzeptor) dotiert ist. Die zahlreichen, sehr scharfen Linien werden durch Elektronenübergänge von einem Donator zum nächstgelegenen Akzeptor hervorgerufen (Donator-Akzeptor-Paarspektrum).
Die Energie $h\nu$, die bei der Rekombination des Elektrons vom Donator mit dem Loch vom Akzeptor frei wird, beträgt $h\nu = E_g - (E_A + E_D) + e^2/(4\pi\varepsilon\varepsilon_0 R)$ (E_A, E_D = Bindungsenergien am Akzeptor bzw. Donator, e = Elektronenladung, ε = statische Dielektrizitätskonstante, ε_0 = elektrische Feldkonstante, R = Abstand Donator–Akzeptor, E_g = Breite der verbotenen Zone). Das Auftreten diskreter Linien beweist, daß die beteiligten Störstellen substitutionell (auf Gitterplätzen) eingebaut sind, da der Abstand nur diskrete Werte annimmt. Obige Gleichung gestattet eine Präzisionsbestimmung von Bindungsenergien sowie die genaue Messung von ε.
Die mit S bzw. N bezeichneten Linien in Abb. 6.3.8–11 sind keine Paarlinien; sie entstehen bei der strahlenden Rekombination eines an den neutralen Donator S bzw. an die isoelektronische Störstelle N gebundenen Excitons. Diese Vernichtungsstrahlung gebundener Excitonen ist der physikalische Grundprozeß der Lichterzeugung in den *lichtemittierenden Dioden* (LED), die z. B. als optische Anzeigelemente in Taschenrechnern verwendet werden.
Das Vorhandensein einer ungeraden Anzahl von Elektronen an einer Störstelle läßt sich mit der Methode der *paramagnetischen Elektronenresonanz* (EPR) nachweisen. Die Wechselwirkung des Elektronenspins mit dem Kernspin des Störatoms gestattet es dabei, die chemische Natur der Störstelle zu identifizieren. Wechselwirkungen mit den Kernspins von Nachbaratomen führen zu sehr linienreichen Spektren, aus denen auf Symmetrie und Einbauort der Störstelle geschlossen werden kann.

6.4. Gitterschwingungen

Die Ionen führen um ihre Gleichgewichtslagen im Festkörper infolge der Nullpunktsenergie und der Temperaturbewegung Schwingungen aus. Auch können durch inelastische Prozesse, z. B. inelastische Neutronenstreuung, Schwingungen angeregt werden.
Infolge des großen Massenunterschiedes zwischen Elektronen und Ionen ist die *adiabatische Näherung* gerechtfertigt: Das Elektronenspektrum $E(k, \boldsymbol{R})$ wird bei festgehaltenen Ionenlagen $\boldsymbol{R}$ berechnet. Die Gesamtenergie des Festkörpers $E_{ges}(\boldsymbol{R})$ kann dann als Funktion der Positionen $\boldsymbol{R}$ betrachtet werden. Die Gleichgewichtslagen $\boldsymbol{R}^0$ ergeben sich (bei $T = 0$) aus der Forderung, daß die Gesamtenergie ein absolutes Minimum aufweist. Die Änderung von $E_{ges}(\boldsymbol{R})$ bei Auslenkungen aus den Gleichgewichtslagen bestimmt die rücktreibenden Kräfte, die ein Ion erfährt, und damit die für die Berechnung des Schwingungsspektrums wesentlichen Kraftkonstanten.

6.4.1. Oszillatorkette

Die Oszillatorkette ist ein eindimensionales Modell von Atomen, die um ihre Gleichgewichtslage schwingen. Es erlaubt die Erläuterung der wesentlichen Begriffe, die beim Schwingungsspektrum von Festkörpern auftreten. Abb. 6.4.1–1a zeigt eine Kette, bei der sich in jeder Zelle der Länge a ein Atom der Masse m befindet. Die (kleinen) Auslenkungen aus den Gleichgewichtslagen werden mit $y_n(t)$ bezeichnet. Bei der Wahl *periodischer Randbedingungen* ist $y_0 = y_N$. Die rücktreibenden Kräfte werden durch Federn mit den Kraftkonstanten $C \equiv m\omega_0^2$ erfaßt.
Die Hamilton-Funktion der Kette ist $H = \sum_{n=1}^{N} \frac{p_n^2}{2m} + \sum_{n=1}^{N} \frac{C}{2}(y_n - y_{n-1})^2$. Es ist eine Eigenschaft des harmoni-



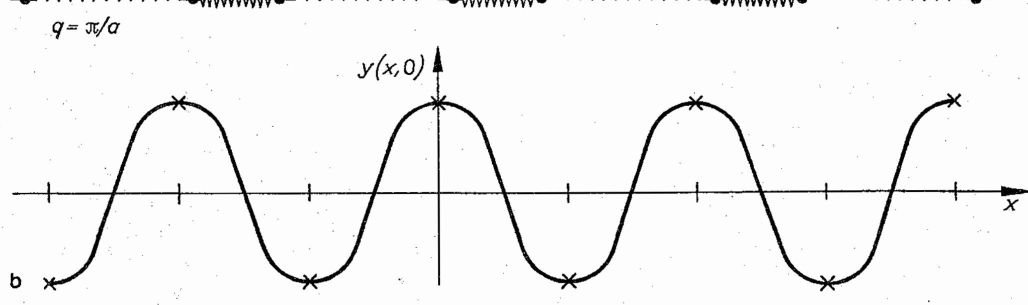


Abb. 6.4.1–1 (a) einatomige Oszillatorkette, (b) Beispiel für ein Schwingungsbild

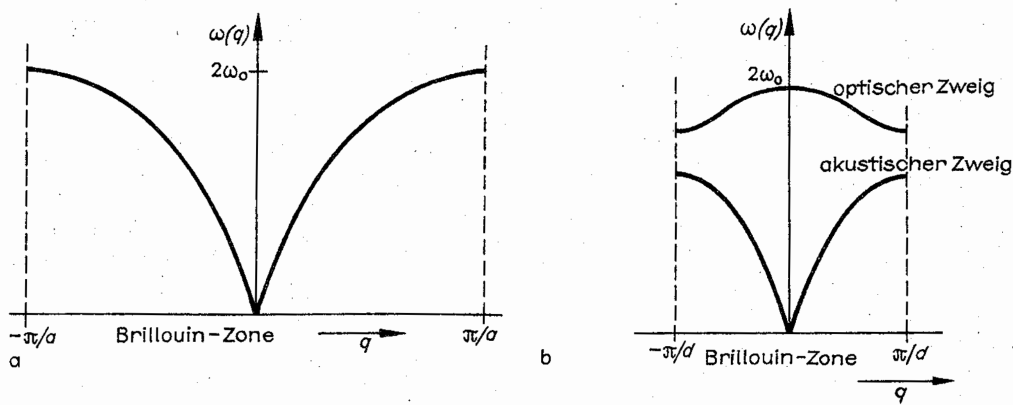


Abb. 6.4.1–2 (a) Schwingungsspektrum einer einatomigen Kette und (b) einer zweiatomigen Kette

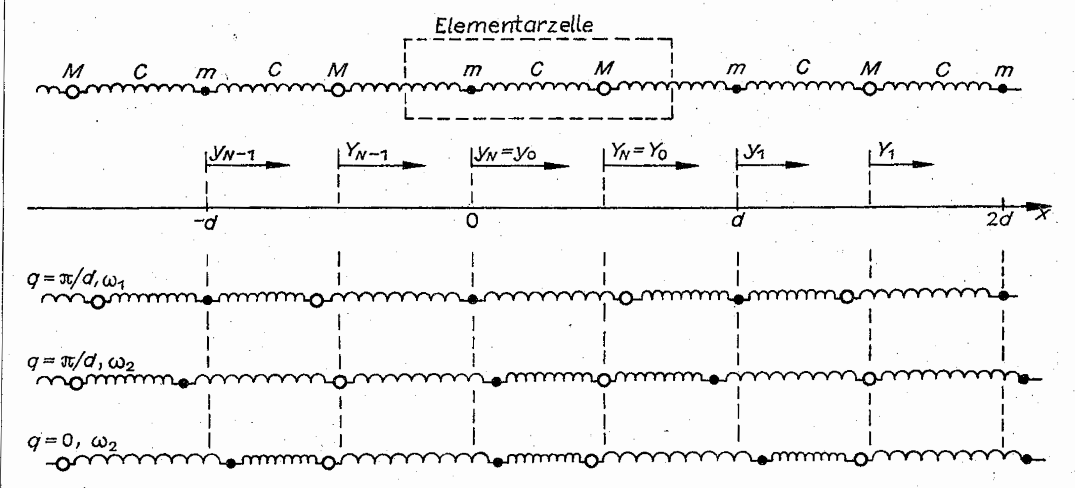


Abb. 6.4.1–3 Zweiatomige Oszillatorkette

schen Potentials, daß sich durch Einführung von Normalkoordinaten $Q_n = \sum_{n'} \alpha_{nn'} y_{n'}$ die Bewegungsgleichungen entkoppeln lassen. Das heißt, die Lösungen der Bewegungsgleichungen sind *Eigenschwingungen*, bei denen alle Massen mit der gleichen Frequenz schwingen, $y_n = A_n\, e^{-i\omega t}$. Bei einer periodischen Kette gilt wegen der Translationsinvarianz das *Bloch-Theorem*. Es sagt hier aus, daß sich die Auslenkungen in benachbarten Zellen nur um einen Phasenfaktor unterscheiden, $A_{n+1}/A_n = A = e^{iqa}$. Physikalisch verschiedene Zustände kann man mit den q-Werten aus der ersten *Brillouin-Zone* $-\pi/a < q \leq \pi/a$ beschreiben. Bei N Elementarzellen sind dies N äquidistante q-Werte. Zu den verschiedenen Schwingungszuständen gehören unterschiedliche Eigenfrequenzen $\omega(q)$.

6.4.1.1. Eindimensionale einatomige Kette. Für sie ist das Schwingungsspektrum durch $\omega(q) = 2\omega_0\,|\sin(qa/2)|$ (Abb. 6.4.1–2a) gegeben. Führt man statt des Laufindexes n die Variable $x = an$ ein, dann sind die Auslenkungen $y(x, t) = y_0 \cos(qx - \omega t)$. Die zu $q = \pi/a$ gehörende Schwingung ist für den Zeitpunkt $t = 0$ in Abb. 6.4.1–1b veranschaulicht. Bis jetzt wurden nur *longitudinale Schwingungen* diskutiert. Es sind ebenso *transversale* denkbar, bei denen die Auslenkung senkrecht zur Kettenrichtung erfolgt. Es gibt zwei unabhängige Polarisationsrichtungen für transversale Schwingungen.

6.4.1.2 Zweiatomige Kette. Die Bezeichnung in Abb. 6.4.1–3 wurde so gewählt, daß für $m = M$ und $d = 2a$ die in Abb. 6.4.1–1 dargestellte einatomige Kette (mit $2N$ statt N) entsteht. Bei den Eigenschwingungen schwingen wieder alle Massen mit gleicher Frequenz. Abb. 6.4.1–5 enthält einige Schwingungsbilder. Zu jedem q-Wert $(-\pi/d < q \leq \pi/d)$ gibt es zwei Eigenfrequenzen

$$\omega_{1,2}^2(q) = 2\omega_0^2\left[1 \pm \sqrt{1 - \frac{4Mm}{(M+m)^2}\sin^2(qd/2)}\,\right]$$

(Abb. 6.4.1–2b).

Den oberen Zweig nennt man den *optischen* Zweig. Für $q = 0$ schwingen die beiden Atome einer Zelle in Gegenphase, so daß die Schwingung bei Ionenkristallen optisch aktiv ist. Den unteren Zweig nennt man den *akustischen Zweig*. Bei kleinem q, d. h. großen Wellenlängen, entsprechen diese Schwingungen den Schallwellen in Einkristallen.

6.4.2. Phononenspektren

Bei den Schwingungen eines Festkörpers sind die Auslenkungen der Atome durch $\boldsymbol{u} = \boldsymbol{u}_0(\boldsymbol{q}) \cos(\boldsymbol{qr} - \omega t)$ gegeben. $\boldsymbol{q}$ ist ein in der Brillouin-Zone liegender Wellenvektor. Bei einem Atom je Elementarzelle gibt es zu einem $\boldsymbol{q}$ drei Eigenfrequenzen. Das Schwingungsspektrum besteht aus *einem longitudinalen* und *zwei transversalen Zweigen*. Die *Polarisation* der Eigenschwingung ist durch $\boldsymbol{u}_0(\boldsymbol{q})$ gegeben. Bei einer transversalen Schwingung steht $\boldsymbol{u}_0$ senkrecht auf der Ausbreitungsrichtung $\boldsymbol{q}$. Wegen der Anisotropie der Kristalle liegt der Polarisationsvektor jedoch i. allg. nicht parallel oder senkrecht zu $\boldsymbol{q}$.

Hat der Kristall M Atome je Elementarzelle, dann gibt es $3M$ Schwingungszweige $\omega_\nu(\boldsymbol{q})$. Zu jedem Zweig gehören N Eigenschwingungen, d. h. N $\boldsymbol{q}$-Punkte in der Brillouin-Zone. Insgesamt gibt es also $3MN$ Eigenschwingungen, so viele wie Freiheitsgrade bei MN Atomen. Infolge der erforderlichen quantenmechanischen Beschreibung hat jede Eigenschwingung nur die *diskreten Energiewerte* $E_q = \hbar\omega(\boldsymbol{q})(n_q + 1/2)$. Die Energie einer Eigenschwingung kann sich nur um Quanten $\hbar\omega(\boldsymbol{q})$ ändern. Eine Elementaranregung nennt man ein *Phonon*. Zu ihr gehört die Energie $\hbar\omega(\boldsymbol{q})$ und der Quasiimpuls $\boldsymbol{q}$. n_q ist dann die Zahl der zu einer bestimmten Eigenschwingung angeregten Phononen. Phononenspektren lassen sich am besten durch inelastische Neutronenstreuung direkt ausmessen. Aus dem Energieverlust $\Delta E = \hbar\omega(\boldsymbol{q})$ und der Impulsänderung $\Delta \boldsymbol{k} = \boldsymbol{q} + \boldsymbol{G}$ der Neutronen läßt sich $\omega(\boldsymbol{q})$ gewinnen (Abb. 6.4.2–1).

Abb. 6.4.2–2 zeigt die *Phononenzustandsdichte* (vgl. 6.3.7.)

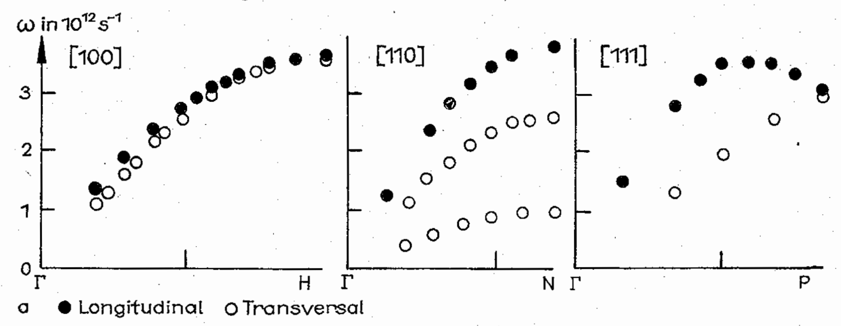
$$D(\omega) = \frac{1}{\Omega_{BZ}} \sum_{\nu}^{\text{Zweige}} \int_{\omega_\nu(\boldsymbol{q}) = \omega} \mathrm{d}^2 q \frac{1}{|\partial\omega_\nu/\partial\boldsymbol{q}|}, \int \mathrm{d}\omega\, D(\omega) = 3M$$

der drei Schwingungszweige von Aluminium.

6.4.2.1. Debye-Modell. Das Debye-Modell für das Schwingungsspektrum ist für qualitative Diskussionen gut geeignet. Aus Abb. 6.4.2–3 ist ersichtlich, daß man auf die Un-

Tab. 6.4.3–1 Debye-Temperatur θ_D einiger Elemente (nach KITTEL)

Element	θ_D in K	Element	θ_D in K
Beryllium	1 440	Kupfer	343
Kohlenstoff	2 230	Molybdän	450
Natrium	158	Palladium	274
Magnesium	400	Silber	225
Aluminium	428	Wolfram	400
Kalium	91	Platin	240
Eisen	470	Gold	165
Nickel	450	Blei	105



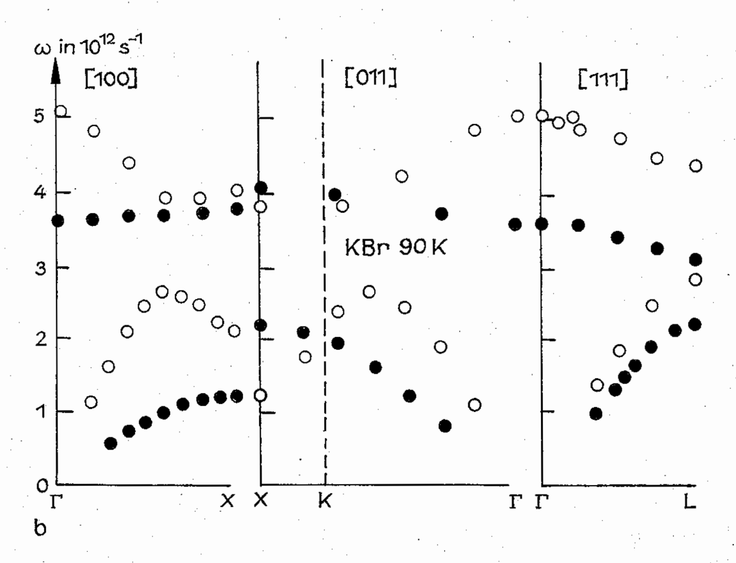


Abb. 6.4.2–1 Phononenspektrum von (a) Natrium und (b) Kaliumbromid (nach WOODS u. a.)

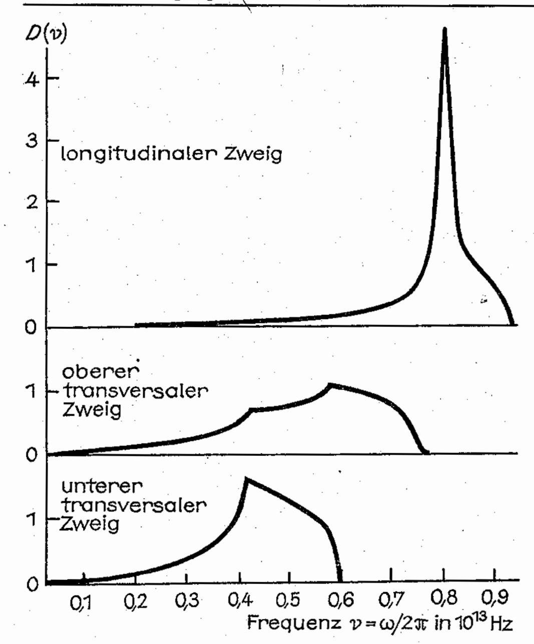


Abb. 6.4.2–2 Phononenzustandsdichte für Aluminium. Van-Hove-Singularitäten sind deutlich zu erkennen (nach WALKER)

terscheidung von optischen und akustischen Zweigen verzichtet. Man betrachtet drei akustische Zweige, deren Frequenzen $\omega(\boldsymbol{q}) = c_S\,|\boldsymbol{q}|$ linear mit der Wellenzahl ansteigen. Die Steigung ist durch die Schallgeschwindigkeit c_S gegeben, wobei man i. allg. noch zwischen longitudinalen ($c_{S,L}$) und transversalen ($c_{S,T}$) Wellen unterscheidet. Die $\boldsymbol{q}$-Werte liegen innerhalb einer Kugel, deren Radius $q_D = (6\pi^2/\Omega_0)^{1/3}$ (Ω_0 = Volumen je Atom) so gewählt wird, daß sie MN Zustände umfaßt. Die höchste Frequenz $\omega_D = c_S\,q_D$ wird als *Debye-Frequenz* bezeichnet. Die Zustandsdichte hat im Debye-Modell die einfache Form $D(\omega) = 3M\,3\omega^2/\omega_D^3$.

6.4.3. Durch das Phononenspektrum bestimmte Meßgrößen

6.4.3.1. Elastische Wellen entsprechen akustischen Schwingungen mit großer Wellenlänge. Aus den Phononenspektren ist zu ersehen, daß die Schwingungsfrequenzen der akustischen Zweige für kleine $\boldsymbol{q}$ proportional zu $\boldsymbol{q}$ sind. Der Proportionalitätsfaktor ist die Schallgeschwindigkeit $c_S = \partial\omega/\partial\boldsymbol{q}$. In der Kontinuumsmechanik ergeben sich elastische Wellen, indem man in der Bewegungsgleichung die Spannungen über die Spannungs-Dehnungs-Beziehung (Hookesches Gesetz) eliminiert. Die Größe der rücktreibenden Kräfte wird dort durch die *elastischen Konstanten* C_{ij} beschrieben. Für einen kubischen Kristall z. B. ist das Verschiebungsfeld $\boldsymbol{u}(\boldsymbol{r}, t)$ aus (ϱ = Massendichte)

$$\varrho\frac{\partial^2 u_x}{\partial t^2} = C_{11}\frac{\partial^2 u_x}{\partial x^2} + C_{44}\left(\frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2}\right) + (C_{12} + C_{44})\left(\frac{\partial^2 u_y}{\partial x\,\partial y} + \frac{\partial^2 u_z}{\partial x\,\partial z}\right)$$

zu berechnen.

Die Ausbreitungsgeschwindigkeiten werden durch die elastischen Konstanten bestimmt (vgl. 2.3.5–3). Es ergibt sich also ein unmittelbarer Zusammenhang zwischen Phononenspektrum und elastischen Konstanten. Bei einem halbphänomenologischen Vorgehen zur Berechnung des Phononenspektrums berechnet man daher die Kraftkonstanten nicht aus den ersten Prinzipien, sondern paßt sie an die experimentell gewonnenen elastischen Konstanten an.

6.4.3.2. Gitteranteil der spezifischen Wärme. Er ist aus der Energie der Gitterschwingungen

$$U(T) = \int \mathrm{d}\omega\, D(\omega)\, \varepsilon(\omega, T),\ \varepsilon(\omega, T) = \hbar\omega\left[\frac{1}{\mathrm{e}^{\hbar\omega/kT} - 1} + \frac{1}{2}\right]$$

zu berechnen. $\varepsilon(\omega, T)$ ist die mittlere Energie eines linearen harmonischen Oszillators im thermodynamischen Gleichgewicht bei der Temperatur T. Die mittlere Anzahl der Phononen einer Eigenschwingung $\varepsilon(\omega, T)/(\hbar\omega) - 1/2$ ist durch die Bose-Einstein-Verteilung gegeben.

Bei einer Auswertung im Debye-Modell führt man die über $k\Theta_D = \hbar\omega_D$ definierte Debye-Temperatur Θ_D ein (Tab. 6.4.3–1). Die Schwingungsenergie wird (bei N Atomen) $U(T) = U_0 + NkT(3\pi4/5)\,(T/\Theta_D)^3$ für $T \ll \Theta_D$, $U(T) = U_0 + 3NkT$ für $T \gg \Theta_D$. Der Gitteranteil der spezifischen Wärme verläuft also bei tiefen Temperaturen proportional zu T^3 (Abb. 6.4.3–2), bei hohen Temperaturen geht er in

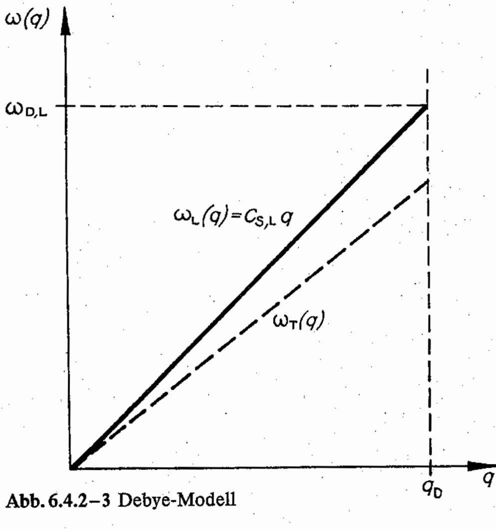


Abb. 6.4.2–3 Debye-Modell

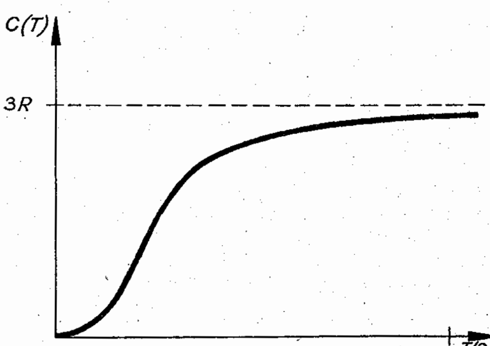


Abb. 6.4.3–2 Gitteranteil der spezifischen Wärme

Übereinstimmung mit der *Dulong-Petitschen Regel* in den konstanten Wert $3R$ ($R = kL$) über, wenn man ein Mol betrachtet, d. h. N gleich der Loschmidtschen Zahl setzt.

6.4.3.3. Wärmeausdehnung. Sie ist eine Folge der Bewegung der Ionen um ihre Gleichgewichtslagen. Sie kann jedoch in der harmonischen Näherung, bei der die potentielle Energie in der Nähe der Gleichgewichtslage durch eine Parabel angenähert wird, nicht gewonnen werden. Erst die *anharmonischen* Anteile im Potential führen dazu, daß bei Anregung die mittlere Position aus dem Minimum herausrückt. Dadurch werden die mittleren Abstände mit wachsender Temperatur größer.

Die *Grüneisen-Formel* folgt aus dem phänomenologischen Bild, daß die rücktreibenden Kräfte mit wachsendem, mittlerem Abstand kleiner werden. Unter der Annahme, daß die relative Änderung aller Frequenzen $\Delta\omega/\omega = -\gamma\,\Delta V/V$ proportional zur Volumenänderung ist, erhält man $\Delta V/V = \gamma\varkappa U(T)$ ($\varkappa$ = Kompressibilität, U = thermische Energie). Der *Wärmeausdehnungskoeffizient*, die Änderung von $\Delta V/V$ mit der Temperatur, wird damit proportional zur spezifischen Wärme. Der Grüneisen-Parameter γ ist dimensionslos und ungefähr von der Größe Zwei.

6.4.4. Einfluß der Gitterschwingungen auf Meßgrößen

6.4.4.1. Beugung von Röntgenstrahlen. In der geometrischen Beugungstheorie muß die Laue-Bedingung $\Delta k = G$ erfüllt sein, damit man einen Beugungsreflex beobachten kann. Bei der Ableitung dieser Bedingung wird angenommen, daß die Ionen an den Gitterplätzen ruhen. Wie wirken sich die Gitterschwingungen auf das Beugungsbild aus? Die Auswertung liefert das Weiterbestehen scharfer Beugungsreflexe in den durch die Laue-Bedingung bestimmten Richtungen. Die Intensität der Beugungspunkte nimmt jedoch ab; sie ist mit dem *Debye-Waller-Faktor* e^{-2W} zu multiplizieren. Dafür tritt ein diffuser Untergrund auf. $2W$ hat die Gestalt

$$2W = 6\frac{(\hbar G)^2/2M}{k\Theta_D}\frac{T}{\Theta_D} \quad \text{für} \quad T \gg \Theta_D,$$

$$2W = \frac{3}{2}\frac{(\hbar G)^2/2M}{k\Theta_D} \quad \text{für} \quad T \ll \Theta_D$$

(M = Ionenmasse, k = Boltzmann-Konstante). Reflexe zu größeren Wellenzahländerungen G werden stärker geschwächt. Die Schwächung für $T = 0$ ist eine Folge der Nullpunktsschwingungen, die auf diesem Wege experimentell nachweisbar sind.

6.4.4.2. Temperaturabhängigkeit des elektrischen Widerstandes. Sie wird wesentlich durch die Gitterschwingungen bestimmt. Qualitativ wächst der Widerstand bei tiefen Temperaturen proportional zu T^5 an, bei hohen proportional zu T, wenngleich er bei vielen Substanzen eine kompliziertere Temperaturabhängigkeit aufweist. Bei Temperaturen, die klein gegen Θ_D sind, sind im wesentlichen nur die Gitterschwingungen bis zu einer Maximalfrequenz $q_{max} = kT$ angeregt. Die zugehörige Wellenzahl q_{max} wächst auch proportional zu T an. Das Volumen der Kugel vom Radius q_{max} im Wellenzahlraum und damit die Zahl der angeregten Phononen ist zu T^3 proportional. Zum anderen ist der Beitrag je streuendem Hindernis auf den elektrischen Widerstand um so größer, je größer der Ablenkwinkel ist. Bewegt sich ein Elektron vor der Streuung in Richtung des elektrischen Feldes mit dem Impuls p und wird es um einen Winkel ϑ abgelenkt, dann ist nach der (elastischen) Streuung die Impulskomponente in Richtung des Feldes $p\cos\vartheta$, sie hat sich um $p(1-\cos\vartheta) \approx p\vartheta^2/2$ geändert. Da der maximale Ablenkwinkel bei kleinem ϑ proportional zu q_{max} ist, wächst der Beitrag je Phonon mit T^2 an. Dies gibt mit der Zunahme der Zahl der Phononen die T^5-Abhängigkeit. Bei hohen Temperaturen sind alle Gitterschwingungen angeregt; q_{max} ist gleich q_D und ändert sich also nicht mehr mit der Temperatur. Der mittlere Beitrag je Phonon wird temperaturunabhängig. Die Anzahl der Phononen wächst proportional mit der Temperatur an.

6.4.5. Elektron-Phonon-Kopplung

In der adiabatischen Näherung werden die Energiebänder $E(\boldsymbol{k})$ für als ruhend betrachtete Ionen berechnet. Will man den Einfluß der Gitterschwingungen auf die Elektronenzustände erfassen, so muß man bedenken, daß ein Elektron eine Gitterschwingung, d. h. ein Phonon, anregen oder absorbieren kann. Die damit verbundenen Energieänderungen sind meist unerheblich, da die Fermi-Energie groß gegen die Phononenenergien ist. Jedoch kann eine wesentliche Impulsübertragung erfolgen. Prozesse dieser Art sind z. B. bei der Berechnung des Idealwiderstandes betrachtet worden. Die Elektron-Phonon-Kopplung wirkt sich auch auf das Quasiteilchenspektrum aus. In Ionenkristallen, wo das Coulomb-Potential eines Elektrons nicht durch Leitungselektronen abgeschirmt werden kann, bewirkt ein in das Leitungsband angeregtes Elektron eine Gitterdeformation in seiner Umgebung, die sich mit ihm durch den Kristall ausbreitet. Das Quasiteilchen Elektron plus Gitterdeformation nennt man ein *Polaron*.

In Metallen tritt eine Beeinflussung der Energiezustände in einer kleinen, durch die Phononenenergien bestimmten Umgebung der Fermi-Energie auf. Insbesondere erhöht sich die Zustandsdichte gegenüber dem aus der Bandberechnung bestimmten Wert $N(E_F) = (1+\lambda)\,N_{Band}(E_F)$. Darin gibt λ den Beitrag der Elektron-Phonon-Kopplung an. Bei genügender Stärke der Elektron-Phonon-Kopplung kann diese zu einer effektiven Anziehung zwischen Elektronen nahe der Fermi-Energie führen. Elektronen mit entgegengesetzter Spinrichtung und entgegengerichtetem Impuls bilden gebundene Zustände, es treten *Cooper-Paare* als Quasiteilchen auf. Im Anregungsspektrum besteht eine Energielücke, die die *Supraleitung* ermöglicht. Die kritische Temperatur für den Übergang in den supraleitenden Zustand ist daher vom Kopplungsparameter der Elektron-Phonon-Kopplung abhängig.

6.5. Realstruktur

Die ideale Struktur eines Kristalles (vgl. 6.1.) ist bei realen Festkörpern wegen der Existenz von Defekten (Kristallbaufehlern) gestört. Es handelt sich dabei nicht um Fehler in dem Sinne, daß auch Kristalle ohne jegliche Abweichung vom idealen Bau herstellbar wären, sondern aus thermodynamischen Gründen enthält prinzipiell jeder gezüchtete Kristall eine Mindestkonzentration von Baufehlern. Der Gleichgewichtszustand des Systems Festkörper bei vorgegebenen Werten von Druck p und Temperatur T ist durch das Minimum der *freien Enthalpie* $G = U - TS - pV$ bestimmt (vgl. 2.6.4.3.), im Gegensatz zur Mechanik, wo das Gleichgewicht dem Minimum der Energie entspricht, da thermi-

sche Vorgänge keine Rolle spielen (U = Energie des Systems und S seine *Entropie*) (vgl. 2.6.4.). Da beim Festkörper auftretende Volumenänderungen klein sind, kann man auch die *freie Energie* $F = U - TS$ betrachten. Der ideale Kristall entspricht dem niedrigst möglichen energetischen Zustand des Systems, d. h. dem Minimum von U. Um Baufehler zu erzeugen, muß Energie dem Festkörper zugeführt werden, damit wird U erhöht. Gleichzeitig erhöht sich aber die Unordnung im Kristall, was eine Erhöhung der Entropie S zur Folge hat. Da eine Verringerung der freien Enthalpie G sowohl durch eine Verkleinerung von U als auch durch eine Vergrößerung von S erreicht werden kann (es ist das negative Vorzeichen bei $-TS$ zu beachten), entspricht der Gleichgewichtszustand, der sich bei einer vorgegebenen Temperatur schließlich einstellt, notwendigerweise einer bestimmten von Null verschiedenen Defektkonzentration. Letztere nimmt mit steigender Temperatur zu, da der Einfluß des Gliedes TS auf die freie Energie und damit der Einfluß der mit weiteren Baufehlern erzeugten Unordnung mit T wächst (vgl. 2.7.4., den Übergang zum idealen Gas ohne Wechselwirkungsenergie der Teilchen, $U = 0$).
Da die meisten Defekte in den Kristallen unter dem Einfluß äußerer Parameter (Temperatur, Druck, mechanische Spannung, elektrische Felder) beweglich sind, bestimmen sie viele makroskopisch meßbare Eigenschaften, die sogenannten *strukturempfindlichen Eigenschaften*. Dazu gehören das mechanische Verhalten bis zum Bruch, die Leitfähigkeit und die dielektrischen Eigenschaften. Dagegen handelt es sich bei der Dichte, der spezifischen Wärme sowie bei der Lichtbrechung um *strukturunempfindliche Eigenschaften*. Während man von Defekten im eigentlichen Sinn als Störung der Nahordnung nur bei kristallinen Festkörpern sprechen kann, existieren auch in amorphen Festkörpern Baugruppen, die bezüglich des makroskopisch beobachteten Verhaltens dieselbe Rolle übernehmen wie die Defekte im Kristall. Allerdings sind deren Eigenschaften wesentlich weniger genau untersucht und bekannt, als das bei Kristallen der Fall ist, so daß sie nur kurz am Beispiel der Gläser skizziert werden können (vgl. 6.5.4.).

6.5.1. Punktdefekte in Kristallen

Bei den Punktdefekten handelt es sich um eine *nulldimensionale* bzw. *atomare Fehlordnung*, d. h. der gestörte Bereich erstreckt sich in allen drei Raumrichtungen nicht weiter als einige wenige Gitterkonstanten (vgl. 6.1.). Zusammenlagerungen von einzelnen Defekten, für die das ebenfalls zutrifft, sollen noch zu den Punktdefekten gerechnet werden, auch wenn die eigentliche Störung mehr als einen Gitterpunkt betrifft. In allen Fällen sind Punktdefekte mit Gitterverzerrungen in ihrer unmittelbaren Nachbarschaft verbunden, da die regulären Gitterteilchen infolge von Relaxationsprozessen in solche Positionen abweichend vom idealen Gitter übergehen, die einem Minimum der Energie entsprechen. Bei der folgenden Charakterisierung der verschiedenen Typen von Punktdefekten werden diese Verzerrungen nicht berücksichtigt, da sie im konkreten Fall außer von der Art des Punktdefektes auch von der Art des Kristalles abhängen.

6.5.1.1. Eigenfehlordnung. Gibt es im Kristall keine *Verunreinigungen* (Fremdteilchen), so kann eine Fehlordnung nur dadurch zustande kommen, daß kristalleigene Teilchen nicht regulär angeordnet sind. Da diese Eigenfehlordnung aus prinzipiellen thermodynamischen Gründen auftritt, spricht man auch von *thermischer Fehlordnung*. Es treten drei Arten der Eigenfehlordnung auf: Leerstellen, Zwischengitterteilchen und Substitutionsfehlstellen.
Leerstellen oder *Lücken* bedeuten, daß ein Teilchen einen im idealen Gitter existierenden Gitterpunkt nicht besetzt (Abb. 6.5.1–1a). Entsprechend dem *Schottky-Mechanismus* kann man sich eine Leerstelle dadurch entstanden denken, daß ein reguläres Gitterteilchen von seinem Platz entfernt und an der Oberfläche auf einem regulären Platz angeordnet wird (Abb. 6.5.1–1b). Durch Minimierung der freien Energie des Kristalles ergibt sich die Anzahl der Leerstellen je Volumeneinheit n_L in einem Kristall, $n_L = N \exp(-E_{BL}/kT)$ (N = Zahl der Gitterplätze je Volumeneinheit, E_{BL} = für die Erzeugung einer Leerstelle erforderliche Bildungsenergie, k = Boltzmannkonstante, T = absolute Temperatur). Für Elementkristalle liegt E_{BL} in der Größenordnung von 1 eV, so daß für die Gleichgewichtskonzentration der Leerstellen n_L/N bei 1000 K etwa 10^{-5}, bei 300 K (Raumtemperatur) etwa 10^{-17} gilt, d. h. bei 1000 K ist etwa jeder 100000-te Gitterplatz, bei Raumtemperatur aber nur jeder 10^{17}-te Platz unbesetzt.

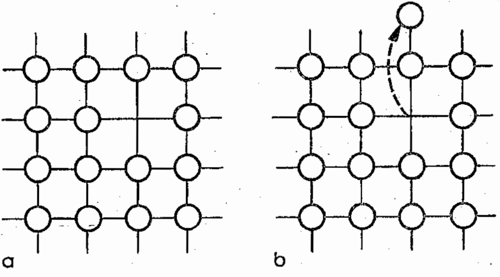


Abb. 6.5.1–1 (a) Leerstelle in einem Elementkristall und (b) ihre Entstehung durch den Schottky-Mechanismus

Zwischengitteratome oder *-ionen* sind zusätzlich in das Gitter auf nichtregulären Plätzen eingebaut. Wegen des relativ großen zusätzlichen Raumbedarfes hat das eine große Verspannung des Gitters und damit eine große Bildungsenergie zur Folge. Bei Metallen liegt sie in der Größenordnung von 3 eV, was bei 1000 K einer Konzentration von nur etwa 10^{-17} entsprechen würde, womit dieser Fehlordnungstyp zu vernachlässigen wäre. Günstigere energetische Verhältnisse liegen jedoch beim *Frenkelmechanismus* vor. Ein Teilchen verläßt einen regulären Gitterplatz und wird im Zwischengitter eingebaut. Die Anzahl der sogenannten Frenkelpaare (Leerstelle und Zwischengitterteilchen) je Volumeneinheit n_F, die thermisch erzeugt werden, ist $n_F = (NN')^{1/2} \exp(-E_{BF}/2kT)$ (N = Zahl der regulären Gitterplätze, N' = Zahl der Zwischengitterplätze je Volumeneinheit, E_{BF} = Bildungsenergie für ein Frenkelpaar).
Substitutionsfehlstellen treten als Eigenfehlstellen in Legierungen auf. In einer geordneten binären Legierung z. B. können Plätze der einen Atomsorte durch Teilchen der zweiten besetzt, d. h. substituiert sein.
Ionenkristalle zeigen insofern Besonderheiten der Eigenfehlordnung, da die Gitterteilchen (Ionen) elektrisch geladen sind. Somit würde durch das Entfernen oder Hinzufügen nur einer Ionenart das Gitter elektrisch aufgeladen werden, was aus energetischen Gründen nicht stattfinden kann. Aus denselben Gründen ist eine Substitutionsfehlordnung nicht möglich. Bei dem Vorhandensein einer *Schottky-Fehlord-*

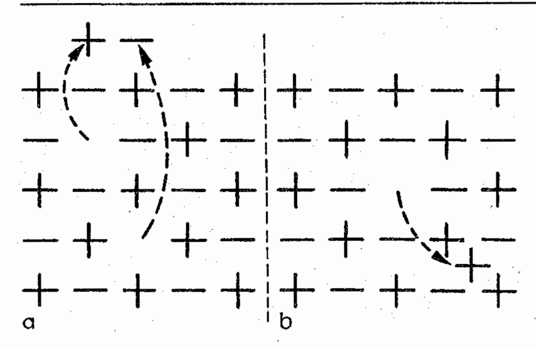


Abb. 6.5.1–2 (a) Schottky-Paar und (b) Frenkel-Paar im Kationenteilgitter eines Ionenkristalles

nung treten deshalb die Leerstellen immer in beiden Teilgittern paarweise auf (Abb. 6.5.1–2a), während bei der *Frenkel-Fehlordnung* die elektrische Neutralität auch in Ionenkristallen automatisch gewährleistet ist, wenn ein Frenkelpaar nur in einem Teilgitter gebildet wird (Abb. 6.5.1–2b). Eine Leerstelle im Kationenteilgitter hat eine negative Überschußladung gegenüber der Umgebung, eine Leerstelle im Anionenteilgitter eine postive. Schottkydefekte werden z. B. in Alkalihalogenidkristallen (z. B. NaCl, KCl), Frenkeldefekte in Silberhalogenidkristallen (z. B. AgCl, AgBr) beobachtet. Die Zahl der Schottkypaare je Volumeneinheit n_S ist durch $n_S = N \exp(-E_{BS}/2kT)$ gegeben, während für die Zahl der Frenkelpaare unverändert der oben angegebene Ausdruck gilt. Es sind N die Zahl der regulären Gitterplätze je Volumeneinheit in einem Teilgitter und E_{BS} die Bildungsenergie für ein Schottkypaar, die in der Größenordnung von 2 eV liegt, so daß für die Konzentrationen die entsprechenden Werte wie für die Leerstellen gelten.

6.5.1.2. Fremdfehlordnung. *Fremdatome* oder *-ionen* sind in allen Kristallen unvermeidlich als Verunreinigung in sehr geringer Konzentration eingebaut (Größenordnung einige *ppm* [*parts per million*] $\triangleq 10^{-6}$ Fremdteilchen je regulärem Gitterplatz), da eine absolute Reinheit des für die Kristallzüchtung benutzten Ausgangsmateriales nicht zu erreichen ist, oder sie werden bewußt als Dotierungselemente bis zu Konzentrationen von einigen Prozent zugesetzt, um die Eigenschaften des Kristalles in definierter Weise zu beeinflussen. Diese Fremdteilchen werden entweder im Zwischengitter oder substitutionell auf regulären Gitterplätzen als Fremd- oder *Störstellenfehlordnung* eingebaut.

Haben die Fremdatome einen wesentlich kleineren Radius als die Teilchen des Wirtsgitters, so befinden sie sich auf

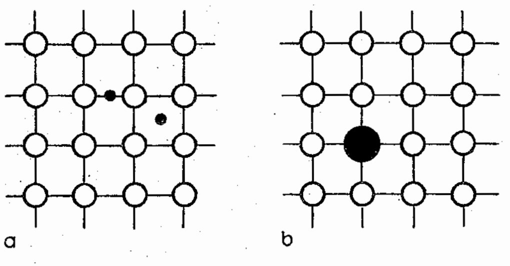


Abb. 6.5.1–3 Einbaumöglichkeiten von (a) Zwischengitterstörstellen und (b) Substitutionsstörstellen

Zwischengitterplätzen (Abb. 6.5.1–3a). Beispiel dafür ist der Einbau von Wasserstoff-, Stickstoff- oder Kohlenstoffatomen im Metallgitter. Wird der Einbau der Fremdteilchen in das Zwischengitter energetisch zu ungünstig, was i. allg. der Fall ist, wenn die Größe der Fremdteilchen mit der der Wirtsgitterteilchen vergleichbar wird, so werden erstere als *Substitutionsfehlstellen* im regulären Gitter eingebaut (Abb. 6.5.1–3b), das gilt z. B. für Kupfer in Aluminium oder Phosphor bzw. Bor in Silicium.

Bei Ionenkristallen ist zu beachten, daß sowohl die Gitterbausteine als auch die Fremdionen elektrisch geladen sind. Das bedeutet zum einen, daß Metallionen jeweils nur in das Kationenteilgitter, Anionen nur in das Anionenteilgitter eingebaut werden, wobei der Einbau vorzugsweise substitutionell erfolgt. Weicht die Ladung der Fremdionen von der der Wirtsgitterionen ab, so muß die elektrische Überschußladung kompensiert werden. Das geschieht i. allg. dadurch, daß zusätzlich Leerstellen entsprechenden Vorzeichens im Gitter entstehen. Beispielsweise wird die positive Überschußladung eines Erdalkaliions (Ca^{++}, Ba^{++}, Sr^{++}) in einem NaCl-Gitter durch eine Kationenleerstelle kompensiert. Wegen der starken anziehenden und relativ weitreichenden elektrostatischen Wechselwirkung zwischen beiden Defekten ordnen sich diese bei Raumtemperatur vorwiegend in benachbarten Gitterpositionen an. Diese sehr stabile Konfiguration wird als *Dipoldefekt* bezeichnet.

6.5.1.3. Farbzentren. Da die Anionenleerstellen in Ionenkristallen positiv geladen sind, wirken sie als Fallen für die Elektronen, die sich im Gitter bewegen können. Solche Elektronen können durch Bestrahlung mit ionisierender Strahlung, z. B. Röntgenstrahlung oder UV-Licht, oder durch Behandlung des Kristalles im Dampf des Wirtsgittermetalles entstehen. Das einfachste und am besten untersuchte Farbzentrum ist das *F-Zentrum* in einem Kristall vom NaCl-Typ, bei dem ein Elektron eine Anionenleerstelle besetzt (Abb. 6.5.1–4). Dieses Elektron verhält sich

+ − + − +
− + − + −
+ − + e^- +
− + − + −

Abb. 6.5.1–4 F-Zentrum im NaCl-Typ-Kristall

im Potential der umgebenden Ionen sehr ähnlich wie ein Elektron im Wasserstoffatom (vgl. 4.3.). Es gibt diskrete Energieniveaus, die das Elektron nach Anregung durch Absorption eines Lichtquantes besetzen kann. Es ist auch möglich, das F-Zentrum zu ionisieren, d. h. das Elektron geht in das Leitungsband (vgl. 6.3.1.) über. Da die Anregungsenergien z. B. bei den Alkalihalogenidkristallen im Bereich des sichtbaren Lichtes liegen, sind die Kristalle bei Anwesenheit von F-Zentren verfärbt, daher der Name Farbzentren. Neben diesem einfach aufgebauten F-Zentrum wurden in Ionenkristallen zahlreiche weitere Farbzentren entdeckt, die teilweise im sichtbaren Bereich, häufig aber auch im ultravioletten oder infraroten Teil des Spektrums absorbieren. Diese Farbzentren entstehen z. B. durch Besetzung einer Anionenleerstelle mit zwei Elektronen, durch Zusammenlagerung von einfachen Zentren zu komplexeren Gebilden und durch Wechselwirkung der Zentren mit in das Wirtsgitter eingebauten Fremdionen. Eine zweite Gruppe von Farbzentren sind die verschiedenartigen V-

Zentren, die durch die Abgabe von Elektronen, d. h. durch die Besetzung mit Defektelektronen, charakterisiert sind. Hier entspricht aber bereits das einfachste bekannte Zentrum einer relativ komplizierten Anordnung der Ionen im Gitter, die nicht näher behandelt werden soll.

6.5.2. Versetzungen im Kristall

Als Versetzungen werden die *eindimensionalen* oder *linienhaften Gitterdefekte* bezeichnet. Sie erstrecken sich in einer Richtung über sehr viele Gitterabstände, während die Störung in den beiden dazu senkrechten Richtungen nach wenigen Gitterkonstanten Entfernung abgeklungen ist. Die innere Energie U des Kristalles wird beim Einbau einer Versetzung so stark vergrößert, daß die mit der Störung des idealen Gitteraufbaues verbundene Entropieerhöhung S diese nicht kompensieren kann (vgl. 6.5.), so daß die freie Energie eines Kristalles mit Versetzungen nicht den minimal möglichen Wert erreicht, d. h. ein Kristall mit Versetzungen befindet sich nicht im thermodynamischen Gleichgewicht. Trotzdem werden wegen meist unvermeidlicher Störungen, wie Temperaturschwankungen, ungleichmäßige Phasengrenzfläche, beim Züchten der Kristalle (vgl. 6.2.) Versetzungen entstehen, die beim Abkühlen eingefroren werden und auch durch nachfolgendes Anlassen bei Temperaturen oberhalb Raumtemperatur nur unvollständig ausgeheilt werden können. Allerdings ist es hier im Gegensatz zu den Punktdefekten (vgl. 6.5.1.) zumindest im Prinzip möglich, versetzungsfreie bzw. versetzungsarme Kristalle oder Kristallbereiche herzustellen. In der Praxis gelungen ist das z. B. für die in der Mikroelektronik als Substrat eine große Rolle spielenden Siliciumeinkristalle.

6.5.2.1. Stufenversetzung. Ein besonders einfacher Versetzungstyp ist die Stufenversetzung. In das reguläre Gitter ist eine Halbebene eingeschoben, die im Kristall endet (Abb. 6.5.2–1), und dadurch Verzerrungen in ihrer unmittelbaren Umgebung verursacht. Das sich linienhaft durch den Kristall hindurchziehende Ende V der eingeschobenen Halbebene wird als *Versetzungslinie* bezeichnet. Da sich in Ionenkristallen eine elektrische Ladung der Versetzungslinie ergeben würde, sind dort immer zwei Halbebenen paarweise derart eingebaut, daß sich deren elektrische Ladung gerade kompensiert.

Die mit den Verformungen in der Nachbarschaft der Versetzung verbundene elastische *Selbstenergie* U_{St} der Stufenversetzung läßt sich mit recht guter Genauigkeit kontinuumsmechanisch, d. h. mit den Methoden der klassischen Elastizitätstheorie, berechnen. Da in unmittelbarer Nähe der Versetzungslinie der atomistische Charakter der Gitterstruktur dominiert, ist eine Kontinuumsnäherung nicht mehr möglich, weil sie zu einer Singularität (Unendlichkeitsstelle) der Energie an der Versetzungslinie führen würde. Dieser Versetzungskern muß deshalb von der Berechnung durch Einführung eines Abschneideradius r_0 (Abb. 6.5.2–2) ausgeschlossen werden. Es ergibt sich dann als Energie der Stufenversetzung

$$U_{St} = (Gb^2/4\pi[1 - \nu])\ln(R/r_0)$$

G = Schubmodul des Kristalles, b = Betrag des Burgersvektors (vgl. 6.5.2.3.), ν = Poissonsche Querkontraktionszahl, R = äußerer Abschneideradius, der sich aus der Größe des betrachteten Kristalles bzw. aus dem mittleren Abstand zweier benachbarter Versetzungen ergibt. Atomistische Berechnungen, bei denen auch die Verschiebungen der einzelnen Gitterbausteine im Versetzungskern berücksichtigt werden, bestätigten die Brauchbarkeit des Ausdruckes für die Selbstenergie.

6.5.2.2. Schraubenversetzung. Ein zweiter Versetzungstyp, der sich anschaulich darstellen läßt, ist die Schraubenversetzung, die man sich durch Verschiebung zweier Kristallbereiche gegeneinander entstanden denken kann, so daß die senkrecht zur Versetzungslinie V verlaufenden Gitterebenen schraubenförmig verbunden sind. Wie bei der Stufenversetzung läßt sich auch für die Schraubenversetzung die *Selbstenergie* U_{Sch} kontinuumsmechanisch berechnen, $U_{Sch} = (Gb^2/4\pi)\ln(R/r_0)$, wobei die Symbole dieselbe Bedeutung wie in 6.5.2.1. haben.

6.5.2.3. Burgers-Umlauf. Der Typ und die Stärke einer Versetzung kann durch die Angabe des zur Versetzung gehörigen *Burgers-Vektors* $\boldsymbol{b}$ charakterisiert werden. Dieser ergibt sich dadurch, daß ein geschlossener Umlauf, der Burgers-Umlauf, um den durch die Versetzung gestörten Kristallbereich vorgenommen wird, wobei der Umlauf die Versetzungslinie V enthalten muß. Parallel zu diesem Umlauf wird ein analoger in einem vollständig ungestörten Bereich durchgeführt, wobei die Zahl der Gitterabstände, die in einer bestimmten Richtung durchlaufen werden, im ungestörten und im gestörten Gitter gleich sind. Die Änderung

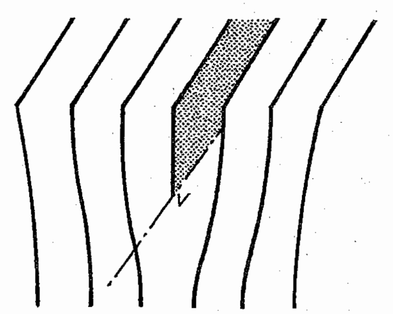


Abb. 6.5.2–1 Stufenversetzung

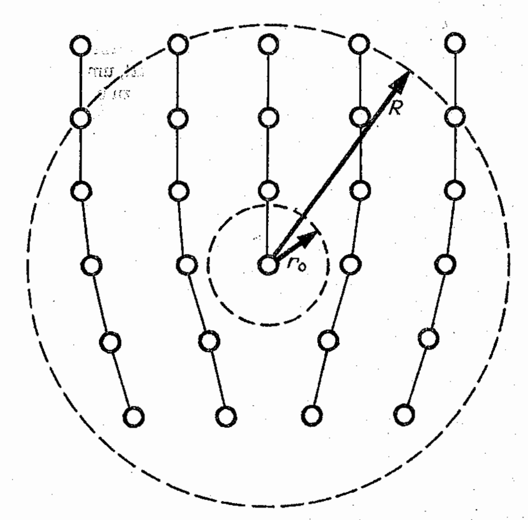
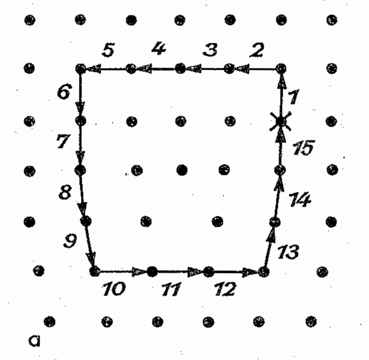


Abb. 6.5.2–2 Zur Berechnung der Selbstenergie einer Stufenversetzung



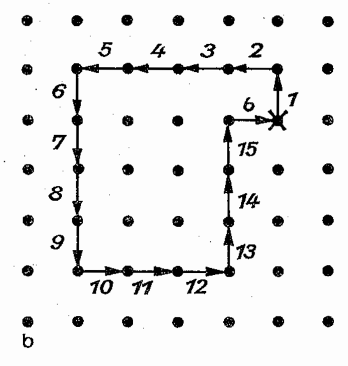


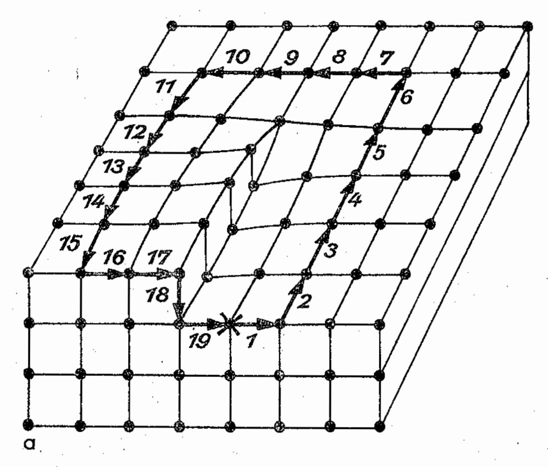
Abb. 6.5.2–3 Burgers-Umlauf um eine Stufenversetzung, (a) gestörter und (b) ungestörter Kristallbereich

der Gitterabstände in der Nachbarschaft der Versetzung wird nicht berücksichtigt, es wird nur die Zahl der Schritte im Umlauf gezählt. Ist der Umlauf um den gestörten Kristallbereich geschlossen, so ist er es i. allg. im ungestörten Bereich noch nicht. Der Vektor, der nötig ist, um auch im ungestörten Gitter den Umlauf zu schließen, ist der Burgers-Vektor, dessen Betrag somit immer durch die Gitterkonstante des idealen Gitters ausgedrückt wird. Der Umlaufsinn wird in der Literatur nicht einheitlich gewählt. Im folgenden werde wie in der Mathematik üblich der Gegenuhrzeigersinn als positiver Umlaufsinn benutzt.

In der Abb. 6.5.2–3 ist als Beispiel der Burgers-Umlauf um eine Stufenversetzung gezeigt. Ausgehend von einem gemeinsamen Ausgangspunkt (markiert durch ein Kreuz) erfolgt der erste Schritt jeweils längs des Vektors *1*. Der weitere Umlauf wird entsprechend den gleichartig numerierten Vektoren durchgeführt. So stehen z. B. die beiden Vektoren *5* und *6* sowohl beim Umlauf um den gestörten als auch um den ungestörten Kristallbereich in der angezeigten Weise nahezu bzw. exakt senkrecht aufeinander. Nach Durchlaufen von 15 Schritten ist beim gestörten Kristallbereich der Umlauf geschlossen, während im ungestörten Bereich noch der Vektor $\boldsymbol{b}$ notwendig ist, um den Umlauf zu schließen. Der Betrag dieses Burgers-Vektors $\boldsymbol{b}$ ist im angegebenen Beispiel gerade gleich einer Gitterkonstanten. In der Abb. 6.5.2–4 ist in analoger Weise ein Burgers-Umlauf um eine Schraubenversetzung skizziert, wobei einander auch wieder gleich numerierte Schritte in beiden Kristallbereichen entsprechen. Hier führt der Burgers-Umlauf um den gestörten Bereich und damit auch zwangsläufig der um den ungestörten Bereich aus der betrachteten Gitterebene heraus, beim Übergang vom Vektor *17* zum Vektor *18*. Um im ungestörten Kristall den Umlauf nach 19 Schritten zu schließen, ist wiederum ein zusätzlicher Vektor $\boldsymbol{b}$ erforderlich, dessen Betrag hier wiederum gleich einer Gitterkonstanten ist. Allgemein ist im kubisch-primitiven Gitter der Betrag des kürzesten Burgers-Vektors gleich der Gitterkonstanten, während z. B. im kubisch-flächenzentrierten Gitter der Betrag des kürzesten Burgers-Vektors gleich dem $1/\sqrt{2}$-fachen der Gitterkonstanten ist. Bei Stufenversetzungen steht der Burgers-Vektor immer senkrecht zur Versetzungslinie, während bei Schraubenversetzungen Versetzungslinie und Burgers-Vektor immer parallel zueinander verlaufen.

Enthält ein Umlauf mehrere Versetzungen, so ist der Burgers-Vektor $\boldsymbol{b}$ in diesem Fall gleich der vektoriellen Summe der Burgers-Vektoren $\boldsymbol{b}_i$ der einzelnen Versetzungen, $\boldsymbol{b} = \sum \boldsymbol{b}_i$. Haben zwei Stufenversetzungen gleiche Stärke, d. h., der Betrag des Burgers-Vektor jeder Einzelversetzung ist gleich der Gitterkonstanten, aber die Burgers-Vektoren sind entgegengesetzt gerichtet, so wird ihre Summe Null.

6.5.2.4. Allgemeine Versetzungen. Die beiden beschriebenen Versetzungstypen sind nur zwei Grenzfälle einer allgemeinen Versetzung, die sich besonders gut veranschaulichen lassen. Prinzipiell ist eine Versetzung durch zwei



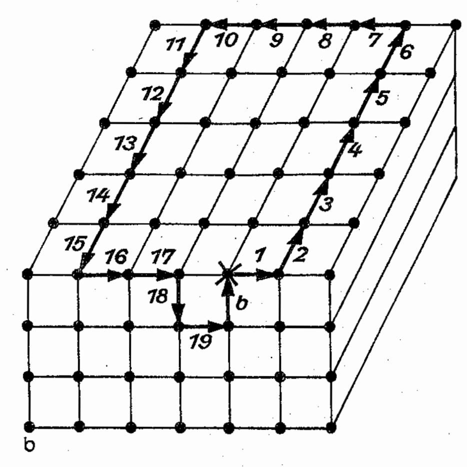


Abb. 6.5.2–4 Burgers-Umlauf um eine Schraubenversetzung, (a) gestörter und (b) ungestörter Kristallbereich

Vektoren vollständig charakterisiert, einen Einheitsvektor s parallel zur Versetzungslinie und den Burgers-Vektor b. Man bezeichnet den Versetzungstyp entsprechend dem Winkel γ zwischen s und b als *γ-Versetzung*, und zwar ist γ durch das skalare Produkt $s \cdot b = b \cos (90° - \gamma) = b \sin \gamma$ definiert (b = Betrag von b), d. h., eine Schraubenversetzung ist eine 90°-Versetzung, während eine Stufenversetzung eine 0°-Versetzung ist.

Für Versetzungen gelten zwei wichtige Sätze. Erstens ist der Burgers-Vektor längs einer unverzweigten Versetzungslinie konstant. Zweitens kann eine Versetzungslinie nie im Inneren des Kristalles enden, sie ist entweder geschlossen oder sie endet an einer Oberfläche, wobei es sich im letzteren Fall auch um innere Oberflächen, z. B. Korngrenzen, handeln kann. Damit kann sich eine reine Stufenversetzung oder eine reine Schraubenversetzung nur von einer Oberfläche zu einer anderen gegenüberliegenden erstrekken, da sie wegen des konstanten Burgers-Vektors nur geradlinig verlaufen können. Das wird nur in Ausnahmefällen zu beobachten sein. Im allgemeinen bilden sich gekrümmte bzw. geschlossene Versetzungslinien im Kristall, bei denen sich der Versetzungstyp von Ort zu Ort ändert und die nur kurze Stücke Stufen- oder Schraubenversetzungen enthalten (Abb. 6.5.2–5), wo an drei Stellen der den Verlauf der Versetzungslinie beschreibende tangentiale Vektor ds eingezeichnet ist.

6.5.2.5. Versetzungsquellen. Während des Kristallwachstums (vgl. 6.2.) entstehen abhängig von den Züchtungsbedingungen Versetzungen im Kristall. Deren Konzentration reicht aber bei weitem nicht aus, die bei Verformungsexperimenten erreichbaren plastischen Verformungen (vgl. 6.6.) zu erklären. Es müssen somit Mechanismen existieren, die eine Versetzungsvervielfachung bewegter Versetzungen bewirken. Zwei solcher Mechanismen wurden bisher ausführlich diskutiert, der Frank-Read-Mechanismus und die Produktion von Versetzungen infolge Quergleitens.

Eine *Frank-Read-Quelle* entsteht dann, wenn auf ein Versetzungsstück, das an beiden Enden fixiert ist, z. B. durch starke Wechselwirkung mit Punktdefekten, eine Schubspannung τ in der Weise wirkt, daß der mittlere Teil des Versetzungsstückes bewegt wird. Das führt zu dem in Abb. 6.5.2–6 für typische Etappen dargestellte Ausbauchen des Versetzungsstückes. Im Stadium *d* ziehen sich die beiden entgegengesetzt orientierten Versetzungssegmente *1* und *2* an, so daß es zur Annihilation dieser beiden Stücke kommt. Daraus resultierend löst sich ein Versetzungsring von der Quelle ab, während zwischen den beiden Fixpunkten ein Versetzungsstück zurückbleibt (Stadium *e*), so daß der Ausgangszustand wieder hergestellt ist und ein weiterer Ring erzeugt werden kann.

Quergleiten eines Versetzungsstückes tritt auf, wenn eine bewegte Versetzung (gestrichelte Linie in Abb. 6.5.2–7) auf ein Hindernis trifft. Ein Versetzungssegment, das Schraubencharakter hat, kann sich auf einer Gleitebene, die mit der angelegten Schubspannung τ einen Winkel bildet, bewegen und folgt erst auf einer benachbarten Gleitebene II, die der ursprünglichen I parallel ist, wieder der alten Bewegungsrichtung. Die beiden sich in der Ebene I bewegenden Segmente bauchen sich unter der Wirkung der Schubspannung τ aus, bis die einander berührenden Segmente sich gegenseitig annihilieren und damit der alte Zustand in dieser Ebene wieder hergestellt ist. Auf der Ebene II dagegen breitet sich das neu entstandene Versetzungssegment aus, so daß jede Quergleitung neue Gleitebenen hervorruft.

6.5.2.6. Beobachtung von Versetzungen. Da Versetzungen

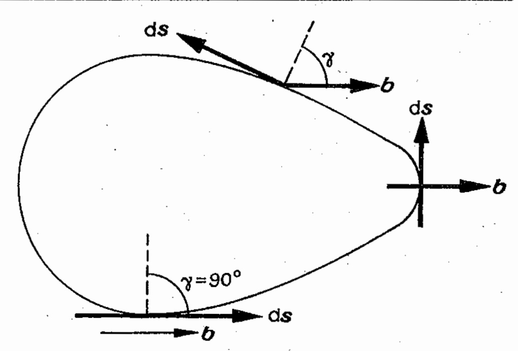


Abb. 6.5.2–5 Geschlossener Versetzungsring

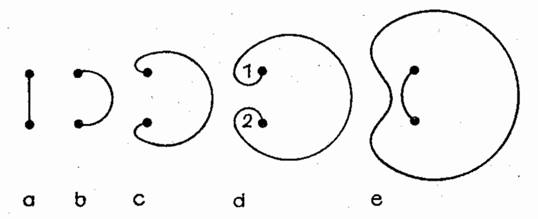


Abb. 6.5.2–6 Stadien der Versetzungsvervielfachung durch Frank-Read-Quelle

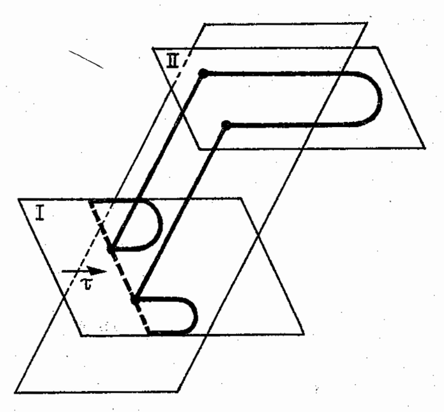


Abb. 6.5.2–7 Versetzungsvervielfachung durch Quergleiten

auf Oberflächen enden, rufen sie dort Gitterverzerrungen hervor. Ätzmittel greifen an diesen energetisch ausgezeichneten Bereichen bevorzugt an, so daß an jedem Durchstoßpunkt einer Versetzung ein *Ätzgrübchen* entsteht, das im Lichtmikroskop beobachtet werden kann. Werden Kristalle im Vakuum bei höheren Temperaturen abgedampft, so schreitet der Materialabbau in diesen gestörten Bereichen besonders schnell voran, so daß nach entsprechender Dekoration eine elektronenmikroskopische Beobachtung möglich ist. In letzterem Fall kann auch die spezifische Struktur aufgelöst werden. Am Durchstoßpunkt von Schraubenversetzungen entstehen z.B. die typischen Abdampfspiralen durch Dekoration der einzelnen Abdampfstufen.

Enthält ein Kristall Fremdatome (vgl. 6.5.1.2.), so werden diese bei geeigneter Temperaturbehandlung bevorzugt an Versetzungen im Kristallinneren ausgeschieden. Man

spricht von einer *Versetzungsdekoration*. Die damit verbundene starke Änderung des Brechungsindex der Umgebung der Versetzung erlaubt eine direkte makroskopische Beobachtung im sichtbaren bzw. infraroten Licht.

Eine direkte *elektronenmikroskopische Beobachtung von Versetzungen* ist bei der Durchstrahlung entsprechend abgedünnter Proben (Dicke in der Größenordnung einiger Zehntel µm bis etwa 1 µm) insbesondere im Höchstspannungselektronenmikroskop (Beschleunigungsspannung 1 MeV, vgl. 7.7) möglich.

Die *röntgengrafische Abbildung von Versetzungen* beruht auf der durch die Versetzungen veränderten Extinktion des Kristallmateriales gegenüber Röntgenstrahlen (vgl. 7.8.).

Außer dem qualitativen Nachweis der Versetzungen erlauben die meisten der angegebenen Verfahren auch, die *Versetzungsdichte* c_V quantitativ anzugeben. Diese ist gleich der Gesamtlänge L der Versetzungslinien je Volumeneinheit $c_V = L/V$, und ist nur näherungsweise gleich der Zahl der Durchstoßpunkte je Flächeneinheit. Ein weiteres Verfahren zur Ermittlung von Versetzungsdichten auch bei stark deformierten Kristallen ist die kernmagnetische Resonanz (vgl. 5.10.). Die Dichte eingewachsener Versetzungen liegt in der Größenordnung von 10^6 bis $10^8\ \mathrm{cm}^{-2}$, während sie infolge plastischer Verformung der Probe noch beträchtlich erhöht werden kann (vgl. 6.6.2.).

6.5.3. Zweidimensionale Defekte in Kristallen

Diese Defekte erstrecken sich flächenhaft im Kristall. Sie lassen sich teilweise aus einer Folge geeignet angeordneter Versetzungen aufbauen.

6.5.3.1. Stapelfehler. Sie entsprechen einer Störung der Schichtenfolge ABABA... bei hexagonal dichtester Pakkung bzw. der Folge ABCABCA... bei kubisch-flächenzentrierter Packung (vgl. 6.1.). Beispielsweise liegt ein Stapelfehler in der fünften Schicht vor, wenn im ersten Fall abweichend von oben die Folge ABABBABA... lautet. Einen Stapelfehler in der siebenten Schicht weist die Folge ABCABCBCA... im kubisch-flächenzentrierten Fall auf.

6.5.3.2. Zwilling. Ein Zwillingskristall bedeutet, daß zwei Teile eines Kristalles spiegelbildlich zueinander angeordnet sind (Abb. 6.5.3–1).

6.5.3.3. Korngrenze. Die Kristallite im polykristallinen Gefüge sind durch *Klein-* bzw. *Großwinkelkorngrenzen* (vgl. 6.1.) gegeneinander abgegrenzt. Kleinwinkelkorngrenzen, bei denen benachbarte Kristallite nur um wenige Grad gegeneinander verkippt sind, entstehen durch in regelmäßigem Abstand hintereinander angeordnete Stufenversetzungen (Abb. 6.5.3–2). Der Verkippungswinkel $\varphi = b/h$ hängt vom Betrag b des Burgers-Vektors der Versetzungen und vom Abstand h der Versetzungslinien ab.

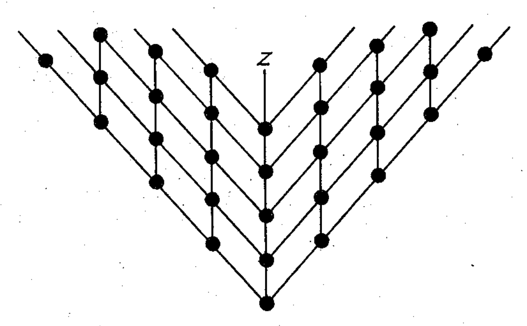


Abb. 6.5.3–1 Kristallzwilling (Z = Zwillingsgrenze)

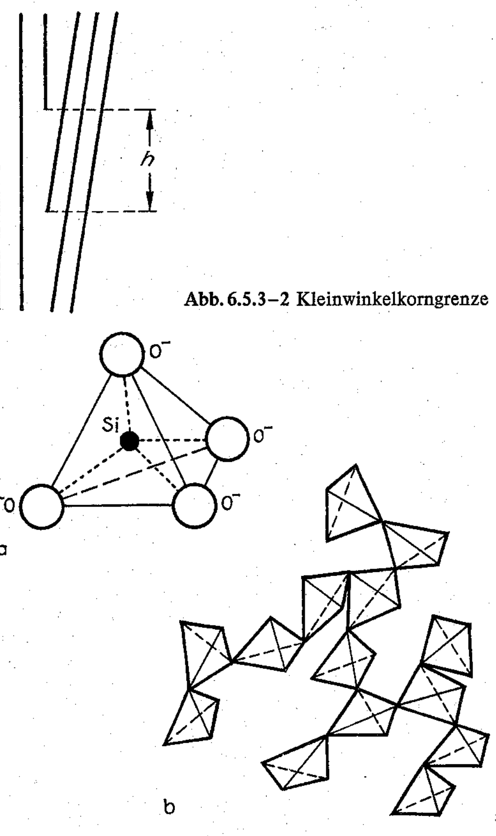


Abb. 6.5.3–2 Kleinwinkelkorngrenze

Abb. 6.5.4–1 (a) Grundbaustein (Tetraeder) und (b) Netzwerk aus Tetraedern im Silikatglas

6.5.4. Glasartige Festkörper

Als typische Vertreter der Substanzen mit nichtkristalliner, d. h. amorpher Struktur, werden die oxidischen Gläser behandelt, während auf amorphe Metalle oder Halbleiter sowie auf Polymere nicht eingegangen wird. Das Grundgerüst der Glasstruktur besteht aus einem unregelmäßig angeordneten Netzwerk des Oxides des *Netzwerkbildners*. Solche sind z. B. Silicium, Bor, Arsen, Germanium. Beim wichtigsten Vertreter, dem Silikatglas, ist das Grundelement ein $SiO_{4/2}$-Tetraeder, dessen Struktur dem in der kristallinen Phase (Quarz) entspricht: Die Ecken des Tetraeders sind von je einem Sauerstoffion besetzt, während sich das Silicium in der Mitte des Tetraeders befindet (Abb. 6.5.4–1a). Das vierwertige Siliciumion sättigt je eine Bindung eines Sauerstoffions ab, so daß der Tetraeder vierfach negativ geladen ist. Über jedes der Sauerstoffionen wird die Verbindung zu einem weiteren Tetraeder hergestellt. Im Unterschied zur Quarzstruktur hat aber der Bindungswinkel zwischen benachbarten Tetraedern keinen konstanten Wert, sondern variiert (Abb. 6.5.4–1b), wodurch die Netzwerkstruktur des Glases zustande kommt. Da die Grund-

struktur der $SiO_{4/2}$-Tetraeder sehr starr ist, wird dieses Gerüst nur schwer von außen angegriffen, was die sehr gute Korrosionsbeständigkeit des reinen SiO_2-Glases (*Kieselglas, Quarzglas*) zur Folge hat.
Eine Auflockerung der Glasstruktur erfolgt durch *Netzwerkwandler*, z. B. durch den Zusatz von Alkali- oder Erdalkalioxiden. Es werden Sauerstoffbrücken der Tetraedergrundstruktur aufgebrochen und die Netzwerkwandler als Ionen in die entstehenden Hohlräume eingebaut. In Abb. 6.5.4–2a wird ein Natriumsilikatglas gezeigt. Die elektrische Ladung des nichtbrückenbildenden Sauerstoffions an der einen Tetraederecke wird durch das Natriumion kompensiert. Da der Netzwerkwandler als Oxid, d. h. Na_2O, in die Glasschmelze eingebracht worden war, bildet das Sauerstoffion des Na_2O ein zweites nichtbrückenbildendes Sauerstoffion, dessen Ladung durch das zweite Natriumion des Oxides kompensiert wird. Beim Einbau eines Erdalkalioxides als Netzwerkwandler neutralisiert das Erdalkaliion die Ladungen zweier benachbarter nichtbrückenbildender Sauerstoffionen (Abb. 6.5.4–2b). Im Gegensatz zu den Netzwerkbildnern sind die Netzwerkwandler im Netzwerk beweglich. Da sie elektrische Ladung tragen, entspricht ihre Rolle der von Punktdefekten in Ionenkristallen. Da sie auch sehr gut mit Wassermolekülen der Atmosphäre reagieren, sind sie für die geringere Korrosionsbeständigkeit dieser Gläser im Vergleich zum Kieselglas verantwortlich.

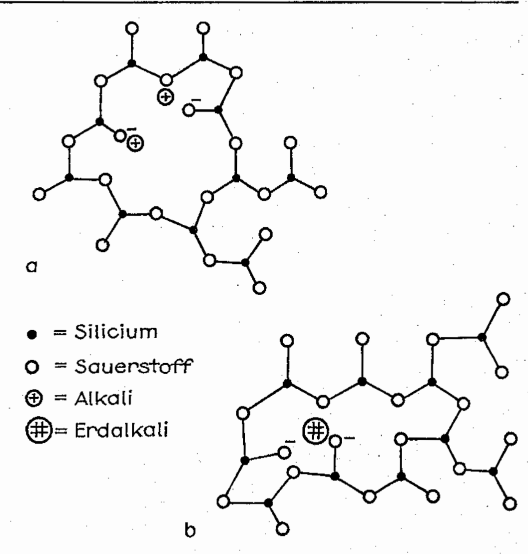


Abb. 6.5.4–2 Ebene Darstellung (a) eines Alkali- und (b) eines Erdalkali-Silikatglases

6.6. Mechanische Eigenschaften

Während in 2.2.1. das mechanische Verhalten der Festkörper rein phänomenologisch, d. h. ohne Berücksichtigung struktureller Einzelheiten behandelt wurde, steht hier der letztere Gesichtspunkt im Vordergrund. Das Konzept der Versetzungen (vgl. 6.5.2.) spielt hier die zentrale Rolle, reicht aber beim derzeitigen Stand der Kenntnis nicht aus, alle Beobachtungen im Rahmen einer geschlossenen Theorie zu deuten.

6.6.1. Elastische Verformung

Wirkt eine mechanische Spannung auf den Festkörper, so bewegen sich die Gitterbausteine in ihrem gegenseitigen Wechselwirkungspotential relativ zueinander. In Abb. 6.6.1–1 ist der Potentialverlauf für ein Teilchen angegeben, dessen Gleichgewichtslage im spannungsfreien Zustand gleich der Position A im Potentialminimum ist. Das Potential ist asymmetrisch, kann aber in der Nähe des Minimums gut durch die gestrichelt angedeutete, zu A symmetrisch liegende Parabel angenähert werden. Kleine äußere Spannungen bewirken deshalb eine Energieänderung ΔE infolge der Verschiebung Δr des Teilchens aus der Gleichgewichtslage, $\Delta E \sim (\Delta r)^2$. Die auf das Teilchen wirkende Kraft F ist $F = -\mathrm{d}\Delta E/\mathrm{d}\Delta r \sim \Delta r$, d. h., es gilt für kleine Verschiebungen das *Hookesche Gesetz* (vgl. 2.2.1.1.). Entlastung hat sofort eine Rückkehr des Teilchens in das Potentialminimum zur Folge, so daß es sich hier um eine rein elastische Verformung handelt.

6.6.2. Plastische Verformung

Größere Spannungen bewirken Verschiebungen der Gitterteilchen in Bereiche, die nicht mehr durch ein symmetrisches Potential beschreibbar sind. Im makroskopischen Verformungsversuch bedeutet das, daß der Zusammenhang zwischen Spannung und Verformung nicht mehr linear ist, aber die Teilchen erreichen bei Entlastung wieder die Gleichgewichtslage, d. h., die Verformung erfolgt unterhalb der Elastizitätsgrenze (vgl. Abb. 2.2.1–1). Ist die Spannung so groß, daß das Teilchen die benachbarte Potentialmulde erreicht, so erfolgt auch bei Entlastung keine Rückkehr in die ursprüngliche Gleichgewichtslage, d. h., der Körper wurde plastisch verformt.

6.6.2.1. Phänomenologische Beschreibung. Die Plastizierung läßt sich durch das Abgleiten paralleler Ebenen gegeneinander beschreiben, wobei diese *Gleitebenen* gegen die Richtung der äußeren Spannung geneigt sind. Der Winkel zwischen Ebenennormale und Spannungsrichtung ist α. Die *Gleitrichtung* schließt mit der Spannungsrichtung den Winkel β ein. Dann wirkt in der Gleitebene die Schubspannung $\tau = \tau \cos\alpha \cos\beta$ (τ = Normalspannung). Der *Orientierungsfaktor* $\mu = \cos\alpha \cos\beta$ kann maximal den Wert $\mu_{max} = 0{,}5$ erreichen (für $\alpha = \beta = 45°$).

6.6.2.2. Theoretische kritische Schubfestigkeit. Damit Abgleiten einsetzt, muß die in der Gleitebene wirkende

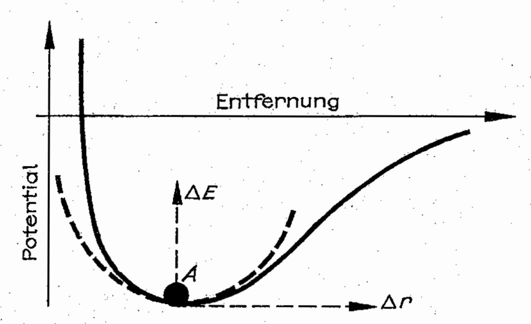


Abb. 6.6.1–1 Potentialverlauf für ein Teilchen im Gitter

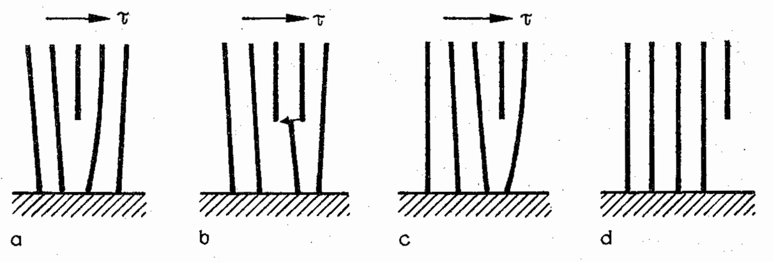


Abb. 6.6.2-1 Bewegung einer Stufenversetzung infolge einer Schubspannung

Schubspannung τ einen kritischen Wert erreichen. Dieser läßt sich theoretisch abschätzen, wenn angenommen wird, daß sich infolge des periodischen Gitterpotentiales die Schubspannung in Abhängigkeit von der Verschiebung x des Teilchens ebenfalls periodisch ändert, $\tau = \tau_{th} \sin(2\pi x/a)$ (a = Gitterkonstante, τ_{th} = theoretische Schubfestigkeit.) Da Abgleiten bereits bei kleinen Verformungen einsetzt, kann τ näherungsweise in eine Reihe entwickelt werden, $\tau = \tau_{th}(2\pi x/a)$. In diesem Bereich gilt aber das Hookesche Gesetz (vgl. 2.2.1.2. für Scherung) $\tau = G\gamma$, mit der Schiebung $\gamma = x/a$, da a der Abstand benachbarter, gegeneinander abgleitender Netzebenen ist. Werden beide Ausdrücke für τ gleichgesetzt, so folgt für die *theoretische kritische Schubfestigkeit* $\tau_{th} = G/2\pi$. Der Vergleich mit Meßergebnissen zeigt jedoch, daß die experimentellen kritischen Schubfestigkeiten mehrere Größenordnungen kleiner als die theoretischen sind, so daß eine plastische Verformung nicht in einer Bewegung von vollständigen Gitterebenen gegeneinander bestehen kann.

6.6.2.3. Elementarmechanismus der Abgleitung. Die plastische Verformung bei relativ kleinen Schubspannungen kann durch Berücksichtigung der Versetzungen (vgl. 6.5.2.) verstanden werden. Enthält z.B. ein Kristall eine Stufenversetzung (Abb. 6.6.2-1a), so bewirkt eine Schubspannung τ, daß sich die eingeschobene Halbebene in Richtung von τ zwar nur geringfügig bewegt, durch Wechselwirkung mit der benachbarten Gleitebene aber zu einer vollständigen Ebene wird, während die benachbarte obere Ebene zur eingeschobenen Halbebene wird (Abb. 6.6.2-1b + c). Damit hat sich die Versetzung tatsächlich um eine Gitterkonstante bewegt, aber nur wenige Gitterteilchen haben geringfügige Verschiebungen erlitten, während der größte Teil der Gitterbausteine in der Gleitebene überhaupt nicht beeinflußt wurde. Deshalb ist auch die erforderliche Schubspannung für diese Bewegung wesentlich geringer als für den Fall, der der Berechnung der theoretischen kritischen Schubfestigkeit zugrunde lag. Ist die Versetzung vollständig durch den Kristall hindurchgelaufen (Abb. 6.6.2-1d), so ist der gesamte obere Kristallbereich gegenüber dem unteren um den Betrag eines Burgers-Vektors (vgl. 6.5.2.3.) abgeglitten. Da als Folge der plastischen Verformung neue Versetzungen entstehen (vgl. 6.5.2.5.), können die Abgleitungen längs einer Gleitebene beträchtliche Werte erreichen, so daß *Gleitstufen* entstehen, die bereits mit dem Lichtmikroskop beobachtbar sind.

6.6.2.4. Verfestigung. Versetzungen können sich bei angelegter Schubspannung nicht ungestört durch den Kristall bewegen. Zum einen wirken Fremdstörstellen (vgl. 6.5.1.2.) bzw. Zusammenlagerungen solcher Punktdefekte als Hindernisse, zum anderen sind unbewegliche Versetzungen (Waldversetzungen) Festhaltepunkte für die bewegten Versetzungen. Da für die Überwindung der Hindernisse eine zusätzliche Kraft aufgebracht werden muß, können durch Dotierung mit Fremdteilchen die mechanischen Eigenschaften, insbesondere die kritische *Fließspannung*, beeinflußt werden.

Wird in einer Gleitebenenschar die Schubfestigkeit erreicht, so tritt Abgleiten in diesem Gleitsystem ein. Im Laufe der Verformung drehen sich Gleitebene und -richtung in Richtung der Zugspannung, d.h. der Orientierungsfaktor (vgl. 6.6.2.1.) wird kleiner. Das hat zur Folge, daß die Schubspannung in einem anderen Gleitsystem solange steigt, bis dort ein Abgleiten einsetzt, so daß sich jetzt mehrere Gleitsysteme schneiden. Dadurch behindern sich die infolge der Versetzungsvervielfachung (vgl. 6.5.2.5.) entstehenden Versetzungen gegenseitig, und es kommt zur *Verfestigung*.

6.6.2.5. Innere Reibung. Der Anteil der Gesamtspannung, der notwendig ist, eine Versetzung im Spannungsfeld der benachbarten Versetzungen zu bewegen, wird als Reibungsspannung bezeichnet. Handelt es sich bei der äußeren Spannung um eine periodisch wechselnde, z.B. infolge von Ultraschall (vgl. 2.3.9.), so werden die sich ausbreitenden Wellen gedämpft. Die Ursache ist die periodische Bewegung der Versetzungssegmente, die an Hindernissen (Aggregate von Punktdefekten, Waldversetzungen) festgehalten werden. Die äußere Spannung reicht nicht aus, die Hindernisse zu überwinden, so daß die Segmente zwischen den Festhaltepunkten ähnlich wie eine gespannte Saite schwingen und infolge der Reibung Energie verlieren. Abhängig vom mittleren Abstand der Hindernisse ergeben sich Eigenfrequenzen der schwingenden Versetzungssegmente. Stimmen Erregerfrequenz und Eigenfrequenz überein (Resonanz), so kommt es zu besonders starken Energieverlusten (*Resonanzabsorption*), so daß in Abhängigkeit von der Frequenz in diesen Bereichen *Dämpfungsmaxima* auftreten, deren Analyse Rückschlüsse auf die Art und Anordnung der im Kristall eingebauten Defekte zuläßt.

6.6.3. Härte

Die Härte als Maß für den Widerstand, den ein Körper einem mechanischen Eingriff entgegensetzt, hat für zahlreiche technische Anwendungen eine große praktische Bedeutung, so daß umfangreiches empirisches Untersuchungsmaterial zur Verfügung steht. Dagegen sind die festkörperphysikalischen Vorgänge, die die Härte eines Materiales bestimmen, noch weitgehend ungeklärt.

Abhängig von der Art der Beanspruchung werden verschiedene Härtearten definiert. Bei der *Schleifhärte* wird die vorgegebene Menge eines Schleifmittels auf dem zu untersuchenden Material solange verschliffen, bis es nicht mehr wirkt, und es wird die Menge des abgeschliffenen Materiales gemessen. Bei der *Bohrhärte* wird die Zahl der Umdrehungen gemessen, die nötig ist, mit einem Diamantbohrer

ein Loch vorgegebener Tiefe zu bohren. Die *Ritzhärte* ist ein Maß für den Widerstand, den ein Festkörper dem Ritzen entgegensetzt. Die *Härteskala nach Mohs* (Tab. 6.6.3–1) erlaubt qualitativ die Ritzhärte zu bestimmen, indem festgestellt wird, welches der Materialien den zu untersuchenden Festkörper gerade noch ritzt. Für die Praxis am wichtigsten ist die *Eindruckhärte*, bei der der Widerstand gegen das Eindringen eines Probekörpers gemessen wird.

Tab. 6.6.3–1 Härteskala nach Mohs

Härtegrad	Mineral
1	Talk
2	Gips
3	Kalkspat
4	Flußspat
5	Apatit
6	Feldspat
7	Quarz
8	Topas
9	Korund
10	Diamant

6.6.3.1. Konventionelle Eindruckhärteverfahren. Als konventionell werden alle die Verfahren bezeichnet, bei denen der erzeugte Eindruck erst nach Entlastung vermessen wird. Die *Eindruckhärte H* ist das Verhältnis aus der Last F, mittels der der Eindruck erzeugt wird, und aus der Oberfläche A des bleibenden Eindruckes, $H = F/A$. Bei der *Brinell-Härte HB* wird eine gehärtete Stahlkugel in das Material gedrückt, so daß mit dem Durchmesser d der zurückbleibenden Kugelkalotte und dem Kugeldurchmesser D für die Härte $HB = 2F(\pi D)\left(D - \sqrt{D^2 - d^2}\right)$ folgt. Die *Knoop-* und die *Vickers-Härte HV* werden durch den Eindruck von Diamantpyramiden gemessen, wobei die Oberfläche A durch Vermessen der Eindruckdiagonalen d bestimmt wird; $H = F(K/d^2)$ (K = Geometriekonstante). Bei der Messung nach Knoop wird eine Pyramide mit rhombischer Grundfläche, bei der am häufigsten angewendeten nach Vickers eine mit quadratischer Grundfläche benutzt, wobei $K = 1{,}854$ ist.

Die konventionell gemessene Eindruckhärte hat zwei wesentliche Nachteile. Einmal tritt durch die teilweise Rückverformung des Eindruckes bei Entlastung infolge der *elastischen Rückfederung* eine Verfälschung der für die Berechnung der Härte verwendeten Oberfläche A des Eindruckes auf, die bei spröden Substanzen besonders groß ist. Zum anderen ist die Härte bei den meisten Materialien lastabhängig, d. h., exakt vergleichbar sind nur Härten, die bei der gleichen Last F gemessen wurden.

6.6.3.2. Registrierende Eindruckhärteverfahren. Zur Vermeidung der Nachteile der konventionellen Verfahren wird die Härte unter Last gemessen, indem während des Eindringens des Probekörpers in das Material sowohl Last F als auch Eindringtiefe p registriert werden; der Zusammenhang läßt sich durch $F = a_1 p + a_2 p^2$ darstellen, wobei die Konstanten a_1 und a_2 lastunabhängig sind. In einer F/p-gegen p-Darstellung ergeben sich Geraden, aus deren Anstieg a_2 nach Fröhlich und Grau eine lastunabhängige Härtezahl $L_2VH = K'a_2$ berechnet werden kann (K' = Geometriekonstante, für Vickers-Pyramide $K' = 0{,}03784$).

Für duktile Materialien, z. B. Metalle, bei denen die elastische Rückfederung vernachlässigbar ist, gilt $HV \approx L_2VH$, für spröde Substanzen $2L_2VH \lesssim HV \lesssim 3L_2VH$, wobei der linke Grenzwert z. B. für Alkalisilikatgläser, ein Zwischenwert für kristalline Keramiken mit Ionenbindung, und der rechte Grenzwert z. B. für Glaskeramiken gemessen wurde. Ein zusätzlicher Vorteil der registrierenden Verfahren gegenüber den konventionellen besteht darin, daß *Oberflächen-* und *Volumenhärte* unterschieden werden können, was besonders für oberflächenveredelte Werkstoffe große Bedeutung hat.

6.6.3.3. Zusammenhang der Härte mit Festigkeitskennwerten. Bei duktilen Materialien ist das Verhältnis zwischen Zugfestigkeit σ_Z und Eindruckhärte H nach Brinell bzw. Vickers konstant, es gilt z. B. für Kohlenstoffstähle $\sigma_Z \approx 0{,}36\,H$ und für legierte Stähle $\sigma_Z \approx 0{,}34\,H$.

Für spröde Materialien wurde eine Beziehung zwischen der *Fließspannung* $\sigma^{(F)}$ (vgl. 2.2.3.1.) und der Eindruckhärte H abgeleitet, $H/\sigma^{(F)} = A + B \ln(E/\sigma^{(F)})$, wobei in Mikrobereichen elastisch-viskoses Verhalten des Materiales vorausgesetzt wird. In dieser Beziehung sind A und B von der Geometrie des Eindruckkörpers abhängige Konstanten und E ist der Elastizitätsmodul. Für eine Vickers-Pyramide gilt $A = 0{,}064$ und $B = 0{,}55$. Die Gleichung erlaubt, aus Härtemessungen die Fließspannung spröder Körper abzuschätzen, was durch makroskopische Experimente i. allg. nicht möglich ist. Für H müssen allerdings registrierend gemessene Werte eingesetzt werden, da bei diesen Substanzgruppen die konventionell ermittelten Werte zu sehr verfälscht sind (vgl. 6.6.3.2.).

6.6.4. Risse

Risse entstehen entweder während der mechanischen Beanspruchung eines Werkstückes oder sie sind bereits vorher vorhanden, z. B. als äußere Beschädigungen an der Oberfläche oder als Inhomogenitäten im Inneren infolge der Herstellung. Während Volumeneigenschaften wie die elastischen Konstanten (vgl. 2.2.1.) oder die Fließgrenze (vgl. 2.2.3.1.) durch Risse kaum beeinflußt werden, hängt die Festigkeit eines Werkstoffes (vgl. 6.6.5.) sehr stark von Art und Anzahl der Risse sowie den Möglichkeiten ihrer Ausbreitung ab.

6.6.4.1. Risse in duktilen Werkstoffen. Die Verformbarkeit duktiler Werkstoffe kommt durch ausreichend bewegliche *Versetzungen* (vgl. 6.5.2.) zustande. Die damit verbundenen plastischen Verformungen bewirken einen Abbau von Spannungsspitzen in der Nachbarschaft vorhandener Risse, so daß Oberflächenrisse nur eine untergeordnete Rolle spielen. Große Verformungen führen jedoch infolge der damit verbundenen *Versetzungsvervielfachung* (vgl. 6.5.2.5.) zum *Aufstau von Versetzungen* an Hindernissen. Dadurch summieren sich die Spannungen zu hohen Werten. Entgegen der abstoßenden Wirkung zweier gleichartiger Versetzun-

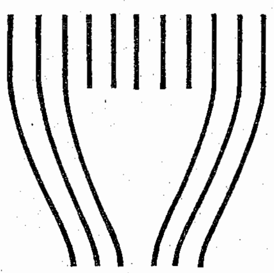


Abb. 6.6.4–1 Bildung eines Anrisses durch Vereinigung mehrerer Stufenversetzungen

gen können sich so mehrere Versetzungen vereinigen und damit einen elementaren Anriß bilden (Abb. 6.6.4–1), der durch Vereinigung mit weiteren Versetzungen im Verlauf der Deformation ständig wächst und somit zur Zerstörung des Werkstückes führen kann.

6.6.4.2. Risse in spröden Werkstoffen. Ein Werkstoff wird dann als spröde bezeichnet, wenn bei makroskopischen Deformationsexperimenten keine merkliche Plastizierung auftritt. Da die plastische Verformung stark temperaturabhängig ist, kann der Begriff *Sprödigkeit* immer nur im Zusammenhang mit einem bestimmten Temperaturbereich für einen Werkstoff verwendet werden, es handelt sich nicht um eine Materialeigenschaft. Dabei zeigen sowohl kristalline als auch amorphe Substanzen sprödes Verhalten, z. B. sind bei Raumtemperatur Germanium und Silikatgläser (Tafelglas) spröde.

Da wegen der fehlenden Plastizierung Spannungen in der Umgebung von Rissen nicht abgebaut werden können, dienen diese infolge der *Kerbwirkung* des Rißgrundes als Spannungskonzentratoren. Für einen elliptischen Riß in einer Platte, dessen große Halbachse l wesentlich größer als die kleine Halbachse ist (Abb. 6.6.4–2), ist die Spannung σ_M im Rißgrund durch $\sigma_M = \sigma_0(1 + 2\sqrt{l/r_k})$ gegeben (r_k = Krümmungsradius der Rißspitze bzw. des Kerbgrundes). Die Spannung nimmt mit der Entfernung von der Rißspitze bis zum Wert σ_0 ab. Für die meist gültige Bedingung $l \gg r_k$ wird $\sigma_M = 2\sigma_0\sqrt{l/r_k}$. Der Riß wird sich dann vergrößern, wenn diese Spannungserhöhung σ_M gleich der *theoretischen Festigkeit* σ_{th} des Werkstoffes ist. Letztere kann nach OROWAN und SMEKAL aus der Arbeit berechnet werden, die erforderlich ist, benachbarte Teilchen soweit voneinander zu entfernen, daß keine Wechselwirkung zwischen ihnen mehr auftritt, $\sigma_{th} = \sqrt{E\gamma/r_0}$ (E = Elastizitätsmodul, γ = Oberflächenspannung, r_0 = charakteristischer mittlerer Abstand der Strukturlemente). Für viele spröde Werkstoffe (Gläser, Keramiken) liefert das die *Orowansche Regel*; $\sigma_{th} \approx 0{,}1\,E$. Aus der

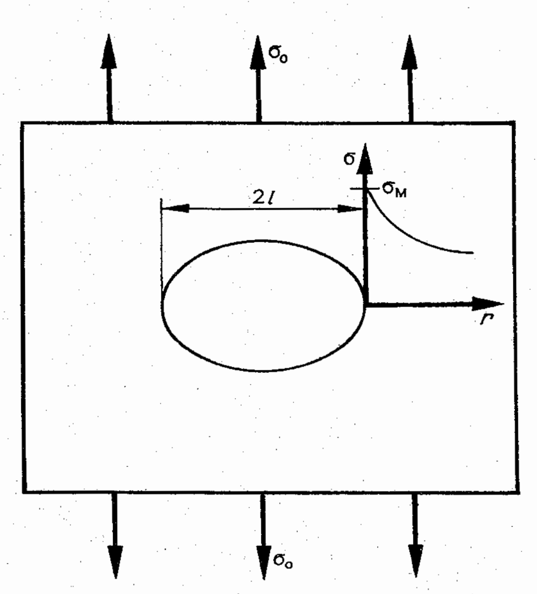


Abb. 6.6.4–2 Spannungserhöhung im Kerbgrund eines Risses

Gleichsetzung von σ_{th} und σ_M folgt für die kritische äußere Spannung σ_{0c}, bei der sich ein Riß der Länge l gerade ausbreitet, $\sigma_{0c} = \sqrt{E\gamma/l}$, wenn $r_k = 4r_0$ angenommen wird. Dieses *Griffithsche Rißkriterium* wurde ursprünglich aus der Bedingung abgeleitet, daß die von der äußeren Spannung geleistete Arbeit ausreicht, die zur Entstehung der neuen Rißoberfläche erforderliche Oberflächenenergie (Oberflächenspannung, vgl. 2.2.4.4.) aufzubringen. Materialeigenschaften sowie die Art des umgebenden Mediums beeinflussen die Rißausbreitung dadurch, daß im Rißgrund Reaktionen zwischen Teilchen aus der Umgebung, z. B. Wassermoleküle, und des Werkstoffes stattfinden, die zur Zerstörung der Bindungen führen. Dabei beeinflussen sowohl die lokalen Spannungen als auch (temperaturabhängige) Diffusionsprozesse die Reaktionsgeschwindigkeit und damit die Rißausbreitung.

6.6.5. Festigkeit

Die makroskopisch meßbare Festigkeit, die *technische Festigkeit* ist i. allg. mehrere Größenordnungen kleiner als die theoretisch berechenbare, die für eine Trennung der Strukturelemente durch die äußere Spannung erhalten wird. Die Ursache dafür sind die bereits vorhandenen bzw. während der Deformation entstehenden Risse. Während erstere für den *Sprödbruch* verantwortlich sind, entstehen letztere beim *duktilen Bruch*. Phänomenologisch unterscheiden sich beide Brucharten dadurch, daß mit dem duktilen Bruch eine starke plastische Verformung des Materiales verbunden ist, es tritt z. B. beim Zugversuch eine große *Bruchdehnung* auf, während der Sprödbruch ohne merkliche plastische Verformung abläuft.

6.6.5.1. Festigkeit bei einsinniger Beanspruchung. Die Richtung der Spannung ändert sich während der Verformung nicht. Spezialfälle sind die *statische Ermüdung* oder der *Kriechversuch*, bei denen die Spannung konstant ist. Bei Versuchsbeginn reicht die Spannung nicht aus, bei den vorhandenen Rissen sofort einen *Gewaltbruch* zu erzeugen. Durch Wechselwirkung mit der äußeren Atmosphäre und durch Diffusionsprozesse im Spannungsfeld vorhandener Risse wird die Rißform, insbesondere der Kerbgrund, so geändert, daß eine Rißausbreitung einsetzt, die zum Bruch führen kann. Bei zügiger Beanspruchung wird die Spannung zeitlich geändert, wobei meist die Dehnungs- oder die Beanspruchungsgeschwindigkeit konstant gehalten werden. Diese Art der Beanspruchung wird manchmal auch als *dynamische Ermüdung* bezeichnet, z. B. in der Glasforschung (vgl. aber 6.6.5.2.), besser wäre sicherlich *quasistatische Ermüdung*. Wegen der den Bruchvorgang beeinflussenden Diffusionsprozesse ist die Festigkeit geschwindigkeitsabhängig. Für Gläser z. B. gilt über mehrere Zehnerpotenzen der Dehnungsgeschwindigkeit v der Zusammenhang $\sigma_B \sim \ln v$ für die Biegebruchfestigkeit.

Während bei spröden Substanzen das *Normalspannungsgesetz* gilt, nach dem der zum *Trennbruch* führende Riß immer senkrecht zur lokal wirkenden Zugspannung verläuft, können bei duktilen Materialien auch *Schiebungsbrüche* durch Schubspannungen auftreten, die immer mit großen plastischen Verformungen verbunden sind.

6.6.5.2. Festigkeit bei wechselnder Beanspruchung. Während der Beanspruchung wechselt das Vorzeichen der Spannung, wobei periodisch wechselnde Spannungen besonders bedeutungsvoll sind (Abb. 6.6.5–1). Mit einer Spannungsamplitude $+\sigma_a$ oszilliert die Spannung um die Mittelspan-

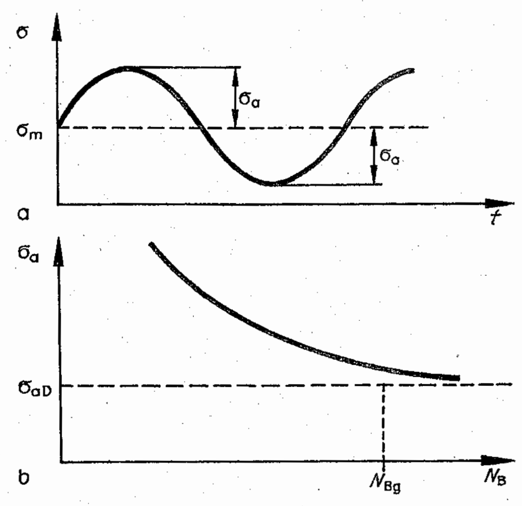


Abb. 6.6.5–1 (a) zeitlich periodisch wechselnde Spannung und (b) Wöhler-Kurve

nung σ_m. Diese Art der Versuchsführung wird i. allg. auch als *dynamische Ermüdung* bezeichnet, z. B. in der Metallforschung (vgl. aber 6.6.5.1.). Für die Festigkeit maßgebend ist die Amplitude, während der Wert der Mittelspannung, solang er ausreichend weit unterhalb der Festigkeit bei einsinniger Belastung liegt, nur sekundäre Bedeutung hat. Bei konstanter Spannungsamplitude σ_a wird die *Bruchlastwechselzahl* N_B gemessen. Die meisten Werkstoffe weisen eine *Dauerfestigkeit* σ_{aD} auf, d. h. für Spannungsamplituden $\sigma_a \leq \sigma_{aD}$ ist die Bruchlastwechselzahl $N_B > N_{Bg}$, wobei N_{Bg} eine vorgegebene *Grenzlastwechselzahl* ist (meist wird $N_{Bg} \geqq 10^6$ gewählt). Mit Spannungsamplituden $\sigma_a > \sigma_{aD}$ befindet man sich im *Zeitfestigkeitsbereich*. Die Ermittlung von *Wöhler-Kurven* (Abb. 6.6.5–1b), die allen Lebensdauerabschätzungen von Bauteilen in der Technik zugrunde liegen, wird einmal dadurch erschwert, daß neben dem Material auch die Form des Bauteiles den Verlauf der Kurve bestimmt. Zum anderen ist die Streuung der Bruchlastwechselzahlen sehr groß, denn der Oberflächenzustand spielt auch bei duktilen Werkstoffen eine große Rolle, da wegen der im Vergleich zur Bruchfestigkeit kleinen Spannung keine merkliche plastische Verformung während des Wechsellastversuches auftritt. Um eine repräsentative Wöhler-Kurve zu erhalten, ist deshalb eine große Anzahl von Einzelversuchen erforderlich.

In vielen Anwendungsfällen ist die Spannungsamplitude nicht konstant, sondern sie wechselt während des Betriebes, z. B. wenn ein Kran bei verschiedenen Lastspielen unterschiedliche Lasten hebt. Wie sich die Aufeinanderfolge verschiedener Spannungsamplituden festkörperphysikalisch auswirkt, ist bisher nicht geklärt. Die sich ergebende *Schadensakkumulation* kann jedoch häufig mit ausreichender Genauigkeit durch die *Miner-Hypothese* $\sum (n_i/N_{Bi}) = 1$ beschrieben werden (n_i = Zahl der Lastwechsel mit der Spannungsamplitude σ_{ai}, N_{Bi} = Bruchlastwechselzahl, die sich aus der Wöhler-Kurve ergibt, wenn das Werkstück nur mit der einzigen Spannungsamplitude σ_{ai} beansprucht worden wäre). Die in Abb. 6.6.5–2 eingezeichnete Stufenfunktion symbolisiert die Zahl der Lastwechsel n_i bei der jeweils dazugehörigen Spannungsamplitude σ_{ai}. Im Punkt B ist gerade die von der Miner-Hypothese geforderte Gleichung erfüllt, und das Werkstück geht zu Bruch. Obwohl auch beträchtliche Abweichungen von der Miner-Hypothese festgestellt worden sind, wird sie wegen ihrer Einfachheit bei praktischen Anwendungen vorwiegend benutzt.

6.7. Elektrische und magnetische Eigenschaften

Wirken elektromagnetische Felder auf einen Festkörper, dann rufen sie Ströme (*Stromdichte* j) und/oder Dipolverteilungen (*Polarisation* P, *Magnetisierung* M) hervor. Diese Reaktionen des Systems sind in der Regel proportional zur äußeren Störung (E, H):

$$j = \sigma E, \quad P = \chi_e \varepsilon_0 E, \quad M = \chi_m H.$$

Die das Material charakterisierenden Größen heißen *elektrische Leitfähigkeit* σ, *elektrische Suszeptibilität* χ_e, und *magnetische Suszeptibilität* χ_m. Treten mehrere Störungen gleichzeitig auf, dann sind erweiterte Zusammenhänge zu betrachten. Ferromagnetika und -elektrika, bei denen auch ohne äußere Felder eine Magnetisierung bzw. Polarisation vorhanden ist, müssen gesondert behandelt werden. Die oben aufgeführten Zusammenhänge sollen hier für verschiedene Typen von Festkörpern diskutiert werden.

6.7.1. Elektronenleitung

Nach ihrer Leitfähigkeit σ bzw. ihrem *spezifischen Widerstand* $\varrho = 1/\sigma$ werden Festkörper in *Metalle, Halbleiter* und *Isolatoren* eingeteilt. Sie unterscheiden sich stark im Zahlenwert von σ und in dessen Temperaturabhängigkeit. Die Leitfähigkeit wird im wesentlichen durch die *Ladungsträgerkonzentration* n, die *effektive Masse* m^* der Ladungsträger und durch deren *mittlere Stoßzeit* τ bestimmt.

6.7.1.1. Metallische Leitfähigkeit. Bei Metallen liegt die Fermi-Energie in einem Energieband. Die Ladungsträgerkonzentration ist groß, 10^{22} bis 10^{23} Elektronen/cm³, sie wird durch die *Zustandsdichte der Elektronen an der Fermi-Energie* $D(E_F)$ (vgl. 4.10.2.) bestimmt. Der spezifische Widerstand liegt bei Zimmertemperatur in der Größe 1 μΩ cm. Der Widerstand steigt in einem weiten Bereich linear mit der Temperatur schwach an, mit wachsender Temperaturbewegung der Gitterbausteine verringert sich die mittlere Stoßzeit. Für Kupfer hat der Temperaturkoeffizient z. B. den Wert $\alpha = (1/\varrho)\,(\partial\varrho/\partial T) = 4 \cdot 10^{-3} K^{-1}$.

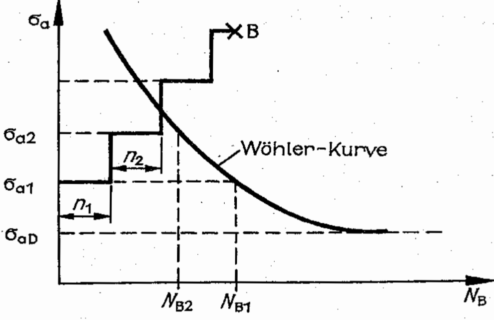


Abb. 6.6.5–2 Schadensakkumulation

In einem qualitativen Bild können die Stöße der Elektronen mit Gitterstörungen im Mittel durch eine geschwindigkeitsproportionale Reibungskraft $\boldsymbol{F}_R = -(m^*/\tau)\dot{\boldsymbol{s}}$ erfaßt werden. Die Bewegungsgleichung $m^*\ddot{\boldsymbol{s}} = e\boldsymbol{E} + \boldsymbol{F}_R$ (vgl. 2.1.3.1.) führt im stationären Fall (verschwindende Beschleunigung $\ddot{\boldsymbol{s}} = 0$) auf die Geschwindigkeit $\dot{\boldsymbol{s}} = (e\tau/m^*)\boldsymbol{E}$, woraus die Stromdichte $\boldsymbol{j} = ne\dot{\boldsymbol{s}} = (ne^2\tau/m^*)\boldsymbol{E}$ folgt. Durch Vergleich mit der Definition $\boldsymbol{j} = \sigma\boldsymbol{E}$ der Leitfähigkeit folgt $\sigma = ne^2\tau/m^*$.

Die konkrete Berechnung der Leitfähigkeit erfordert quantenmechanische Betrachtungen. n (bzw. $D(E_F)$) und m^* folgen aus der Bandstruktur, τ aus dem Wirkungsquerschnitt der Gitterstörungen und der Form der Fermi-Fläche.

6.7.1.2. Supraleitung. Der spezifische Widerstand vieler Metalle verschwindet unterhalb einer *kritischen Temperatur* (Abb. 6.7.1–1a), wie 1911 von KAMERLINGH-ONNES erstmals an Quecksilber beobachtet wurde. T_c liegt bei einigen K, die höchsten bisher beobachteten Sprungtemperaturen bei 23 K. Die Sprungtemperatur wird durch ein Magnetfeld verringert, oberhalb einer *kritischen Feldstärke* $H_c(T) = H_c(0)\,(1-(T/T_c)^2)$ verschwindet die Supraleitung (Abb. 6.7.1–1b). Zur Charakterisierung des supraleitenden Zustandes gehört auch die Beschreibung des *Verhaltens des Supraleiters im Magnetfeld*. Als Meißner-Ochsenfeld-Effekt wird die 1933 gefundene Tatsache bezeichnet, daß das Magnetfeld innerhalb eines Supraleiters – außer in einer dünnen Oberflächenschicht der Dicke $\lambda(T) = \lambda_0[1-(T/T_c)^4]^{-1/2}$ – verschwindet. Das ist ein wesentlicher Unterschied zu einem idealen Leiter, aus $\sigma \to \infty$ folgt nur, daß das Magnetfeld im Leiter eingefroren wird, d. h., $\dot{\boldsymbol{B}} = 0$.

Bei *Supraleitern erster Art* dringt das Magnetfeld nur in eine Oberflächenschicht der Dicke λ ein, in der auch der Suprastrom fließt. Für diese entfällt in Abb. 6.7.1–1b das durch H_{c2} gekennzeichnete Gebiet. Bei *Supraleitern zweiter Art* dringen bei H_{c1} einzelne magnetische Flußfäden in den Supraleiter ein. Der supraleitende Bereich wird mit normalleitenden Schläuchen durchsetzt, bis bei H_{c2} der supraleitende Zustand vollständig verschwindet. Die elektrische Stromdichte $\boldsymbol{j} = \boldsymbol{j}_N + \boldsymbol{j}_S$ ist also i. allg. in einen normalleitenden und einen supraleitenden Anteil zu zerlegen.

Die phänomenologische Beschreibung der Supraleitung erfolgt durch die von LONDON 1935 gefundenen *Materialgleichungen*

$$\frac{\partial}{\partial t}\mu_0\lambda^2\boldsymbol{j}_S = \boldsymbol{E}, \quad \nabla \times \mu_0\lambda^2\boldsymbol{j}_S = -\boldsymbol{B}.$$

Aus den Maxwellschen Gleichungen kann unter Verwendung der London-Gleichungen die Feld- und Stromverteilung bei Anwesenheit von Supraleitern berechnet werden. Die mikroskopische Erklärung der Supraleitung basiert auf der Bildung von *Cooper-Paaren*. Elektronen entgegengesetzter Impuls- und Spinrichtung können in einem durch die Stärke der Elektron-Phonon-Wechselwirkung bestimmten kleinen Bereich um die Fermi-Energie gebundene Paare bilden, die sich wie Bosonen verhalten und bei genügend tiefen Temperaturen ein *Bose-Kondensat* (vgl. 4.10.6.) bilden.

6.7.1.3. Hall-Effekt. An einen Leiter mit Ladungsträgern der Ladung q sei in x-Richtung ein elektrisches Feld $\boldsymbol{E}$ angelegt. Es ruft eine Leitungsstromdichte $\boldsymbol{j} = q n \boldsymbol{v} \parallel \boldsymbol{E}$ hervor. Wirkt auf die Probe ein Magnetfeld $\boldsymbol{B}$ senkrecht zu $\boldsymbol{E}$, in Abb. 6.7.1–2 in z-Richtung, dann wirkt dadurch auf die Ladungsträger die Kraft $q(\boldsymbol{v} \times \boldsymbol{B})$. Sie führt zu Ladungsverschiebungen in y-Richtung, die ihrerseits ein elektrisches Feld $\boldsymbol{E}_H$ aufbauen, bis die magnetische Kraft kompensiert

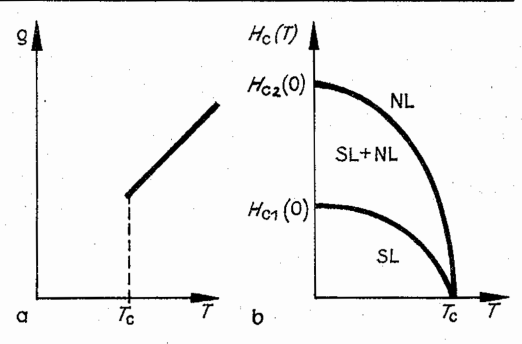


Abb. 6.7.1–1 (a) Abfall des spezifischen Widerstandes eines Supraleiters bei T_0, (b) Abhängigkeit der kritischen Temperatur von der magnetischen Feldstärke H_c. SL = supraleitende Phase, NL = normalleitende Phase

wird, $\boldsymbol{F}_H = q(\boldsymbol{E}_H + \boldsymbol{v} \times \boldsymbol{B}) = 0$. Die in y-Richtung weisende Feldstärke $\boldsymbol{E}_H$ entspricht bei einer Probenbreite b der *Hall-spannung* $U_H = E_H b = IB/(qnd) = R_H IB/d$. d ist die Probendicke und $I = bdj$ der Gesamtstrom. Die Messung des *Hall-Koeffzienten* $R_H = 1/(qn)$ liefert Aussagen über die Ladungsträgerkonzentration und das Vorzeichen der Ladungsträger. Bei Elektronenleitungen ist er negativ. Sind die Ladungsträger Defektelektronen (Löcher, vgl. 6.3.5.), dann ist er positiv.

6.7.1.4. Halbleiter und Isolatoren. Bei Halbleitern und Isolatoren liegt die Fermi-Energie in einer verbotenen Zone. Die Ladungsträgerkonzentration n wird durch die Zahl der thermisch angeregten Ladungsträger bestimmt. Diese Zahl und damit σ wachsen mit der Temperatur proportional zu $\exp[-E_A/(kT)]$ an. Darin ist E_A eine Aktivierungsenergie, z. B. die Breite der verbotenen Zone. Halbleiter und Isolatoren unterscheiden sich aus dieser Sicht nur quantitativ durch die Breite der verbotenen Zone. In Halbleitern liegt n bei Zimmertemperatur in der Größenordnung 10^{15} Ladungsträger/cm^3, der spezifische Widerstand ϱ in einem Bereich von 10^{-2} bis $10^9\,\Omega$cm. Von guten Isolatoren spricht man, wenn $\varrho = 10^{14}$ bis $10^{22}\,\Omega$cm beträgt.

Die Leitfähigkeit von Halbleitern wird durch Fremdatome, Gitterstörungen usw. stark beeinflußt (vgl. 6.3.8.3.). Gerade dies führt zu den für die technische Anwendung so interessanten Eigenschaften.

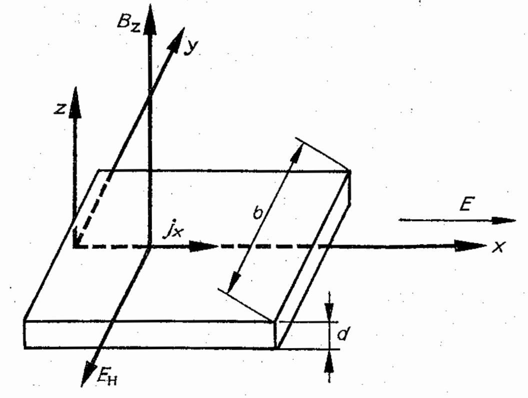


Abb. 6.7.1–2 Hall-Effekt, Richtung E_H für $q < 0$

6.7.1.5. Halbleiter-Leiter-Übergang. Manche Stoffe (z. B. V_2O_3) zeigen in Abhängigkeit von der Temperatur einen Übergang vom Halbleiter zum Leiter, bei dem die Leitfähigkeit bei einer bestimmten Temperatur um mehrere Größenordnungen springt. Dies kann durch eine Veränderung der Bandstruktur durch Änderung des Gitteraufbaus (z. B. eine tetragonale Verzerrung eines kubischen Gitters) oder durch eine Änderung der Stärke der Elektronenkorrelation in Abhängigkeit von der mittleren Elektronendichte (Mott-Übergang) verstanden werden.

6.7.2. Bewegung von Ionen

Infolge von lokalen Fluktuationen der thermischen Energie können sich Gitterbauteilchen – Atome bzw. Ionen – sowie Fremdteilchen im Kristall bewegen, d. h., es existiert eine von der Temperatur T abhängige *Sprungwahrscheinlichkeit* Γ, die Energieschwelle ΔE zu überwinden, um das Teilchen aus seiner Gleichgewichtslage A zu der sogenannten Sattelpunktslage zu bewegen, die zwischen der ursprünglichen und der neuen Gleichgewichtslage B liegt (Abb. 6.7.2–1), $\Gamma = \beta\nu_0 \exp[-\Delta E/(kT)]$ (vgl. 2.6.12.). Hierin sind ν_0 die Schwingungsfrequenz der Teilchen um die Gleichgewichtslage und k die Boltzmann-Konstante. Durch den Faktor β wird berücksichtigt, daß von allen möglichen Bewegungsrichtungen jeweils nur eine für eine bestimmte Bewegung günstig ist. Es ist $\beta = 1/2$ bei ein-, $\beta = 1/4$ bei zwei- und $\beta = 1/6$ bei dreidimensionaler Bewegung. Der makroskopisch beobachtbare Teilchenstrom ergibt sich als Differenz der mikroskopischen Teilchenströme in positiver und in negativer Richtung. Im folgenden wird der Einfachheit halber nur eine eindimensionale Bewegung in x-Richtung betrachtet.

Ein Nettoteilchenstrom $i_x = i_+ - i_-$ in x-Richtung, d.h. ein makroskopisch beobachtbarer Teilchenstrom kommt immer dann im Kristall zustande, wenn entweder ein Konzentrationsgefälle $\partial n/\partial x \neq 0$ existiert oder wenn infolge äußerer Kräfte die Sprungwahrscheinlichkeiten in entgegengesetzten Richtungen $\Gamma_+ \neq \Gamma_-$ voneinander verschieden sind (Abb. 6.7.2–2), $i_x = \alpha\lambda cn(x)(\Gamma_+ - \Gamma_-) - \alpha\lambda^2 c(\partial n/\partial x)(\Gamma_+ + \Gamma_-)/2$ (α = Zahl der Sprungmöglichkeiten, c = mittlere relative Konzentration der im Gitter bzw. im Zwischengitter zur Verfügung stehenden freien Plätze, λ = Abstand zweier Gleichgewichtslagen) (vgl. Abb. 6.7.2–1). Ist die Sprungwahrscheinlichkeit richtungsunabhängig, dann liegt reine Diffusion vor, $i_x = -D\partial n/\partial x$ (vgl. 2.6.11.4.). Der Ausdruck $D = \alpha\lambda^2 c\Gamma$ wird als Diffusionskoeffizient bezeichnet.

6.7.2.1. Ionenleitung. Die Zahl der diffundierenden Teilchen sei ortsunabhängig $\partial n/\partial x = 0$, während eine äußere Kraft F die Sprungwahrscheinlichkeiten beeinflußt. Ist diese Kraft konstant, die potentielle Energie in diesem Kraftfeld also $V(x) = -Fx$ (vgl. 2.1.3.6.), so wird die Energieschwelle in Sprungrichtung um $F\lambda/2$ erniedrigt und entgegen der Sprungrichtung um denselben Betrag erhöht (Abb. 6.7.2–3). Damit sind $\Gamma_+ = \beta\nu_0 \exp[-(\Delta E - F\lambda/2)/kT]$ und $\Gamma_- = \beta\nu_0 \exp[-(\Delta E + F\lambda/2)/kT]$. In den meisten Fällen, sowohl bei einer mechanischen als auch bei einer elektrischen Beanspruchung, ist die äußere Kraft so klein, daß bei Raumtemperatur $F\lambda/2 \ll kT$ gilt, weshalb die Näherung $\Gamma_+ - \Gamma_- = \beta\nu_0(F\lambda/kT)\exp(-\Delta E/kT)$ benutzt werden kann. Handelt es sich um die Bewegung von Teilchen der Ladung q im elektrischen Feld E, d.h., $F = qE$, so wird die elektrische Stromdichte $j = qi_x$ zur Feldstärke E proportional,

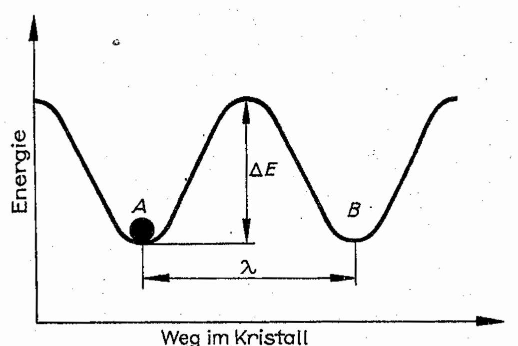


Abb. 6.7.2–1 Thermische Bewegung im Gitter

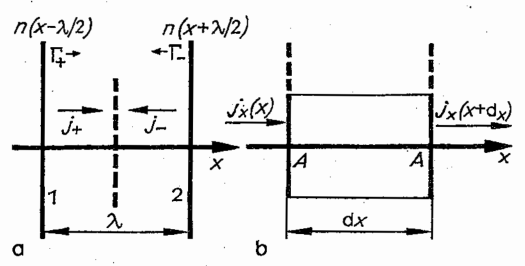


Abb. 6.7.2–2 (a) Teilchenströme und (b) Änderung des Nettoteilchenstromes infolge lokaler Konzentrationsänderungen, $j = qi$

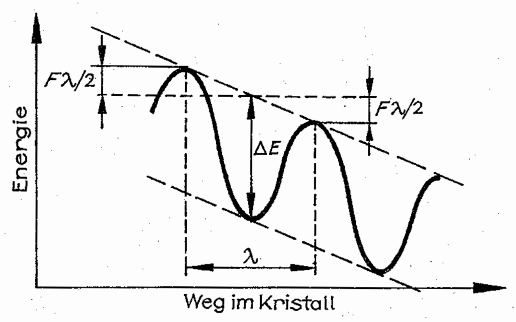


Abb. 6.7.2–3 Bewegung im Kristallgitter bei äußerem angelegtem Feld

d.h., es gilt für $qE\lambda/2 \ll kT$ das *Ohmsche Gesetz für die Ionenleitung* im Kristallgitter. Die Leitfähigkeit hängt in der Form $\sigma \sim (1/T)\exp(-E_L/kT)$ von der Temperatur ab, wobei die Bedeutung der Aktivierungsenergie E_L vom speziellen Mechanismus der Bewegung abhängt.

Das Verhältnis $\sigma/D = nq^2/(kT)$ von Leitfähigkeit zu Diffusionskoeffizient ist die *Nernst-Einstein-Beziehung*, die immer dann gültig ist, wenn Diffusion und Ionenleitung über denselben atomaren Mechanismus ablaufen. Hier ist n die Konzentration der Ladungsträger, im Ionenkristall z.B. entsprechend dem Leitungstyp die Zahl der Kationen bzw. Anionen je Volumeneinheit.

6.7.2.2. Superionenleiter. Verschiedene Ionenkristalle erhöhen ihre Leitfähigkeit in Abhängigkeit von der Temperatur sprunghaft um mehrere Größenordnungen, ohne daß

eine Beteiligung von Elektronenleitung auftritt. σ erreicht Werte, die in Ionenleitern sonst erst oberhalb des Schmelzpunktes beobachtet werden. Man spricht von Superionenleitfähigkeit, wenn σ über $10^{-2}/\Omega$ cm anwächst. In einigen Fällen werden Werte erreicht, die mit der metallischen Leitfähigkeit vergleichbar sind. Lithiumnitrid (Li_3N) hat bei Raumtemperatur eine Leitfähigkeit von $10^3/\Omega$ cm. Eine vielfach untersuchte Substanz ist Silberiodid (AgI), für das σ bei 147 °C um drei Größenordnungen auf $1{,}3/\Omega$ cm anwächst.

Dieses Verhalten kann als Aufschmelzen eines Teilgitters verstanden werden. Alle Ionen dieses Teilgitters – im Falle des AgI die Ag^+-Ionen – nehmen an dem Ladungstransport teil, so daß hohe Ladungsträgerkonzentrationen von $n \approx 10^{22}/\mathrm{cm}^3$ auftreten. Die Ag^+-Ionen befinden sich dabei vorwiegend auf Zwischengitterplätzen, deren Anzahl ein Vielfaches der Zahl der Ag^+-Ionen beträgt. Damit ist auch die Zahl der Sprungmöglichkeiten groß. Die Aktivierungsenergie für einen Sprung in einen anderen Zwischengitterplatz ist relativ klein, wie aus der Beobachtung hervorgeht, daß die Ag^+-Ionen auf den Zwischengitterplätzen stark anharmonische Schwingungen ausführen.

Superionenleiter haben wegen ihres möglichen Einsatzes als Festkörperelektrolyte Interesse gefunden. Man hofft, Batterien hoher Energiedichte, d. h. eines günstigen Verhältnisses von gespeicherter Energie zu Masse der Batterie, bauen zu können.

6.7.3. Polarisation

Unter dem Einfluß eines äußeren elektrischen Feldes $\boldsymbol{E}_a$ werden Festkörper elektrisch polarisiert. Ohne permanente Polarisation bei verschwindendem Feld (vgl. Ferroelektrika) ist die Polarisation bei schwachem elektrischen Feld $\boldsymbol{E}$ proportional zu diesem $\boldsymbol{P} = \chi\varepsilon_0\boldsymbol{E}$, ihre Stärke wird durch die Materialeigenschaft *Suszeptibilität* χ oder *relative Dielektrizitätskonstante* $\varepsilon = 1 + \chi$ gekennzeichnet. Die atomare Polarisierbarkeit charakterisiert das bei einem Gitterbaustein durch das äußere Feld erzeugte elektrische Dipolmoment $\boldsymbol{p} = \alpha\varepsilon_0\boldsymbol{E}_a$. Bei N Gitterbausteinen je Volumen ergibt sich die Polarisation $\boldsymbol{P} = \varkappa\varepsilon_0\boldsymbol{E}_a$ mit der *Polarisierbarkeit* $\varkappa = N\alpha$. Berücksichtigt man, daß zum elektrischen Feld $\boldsymbol{E}$ neben dem äußeren Feld auch das von der Dipolverteilung erzeugte elektrische Feld beiträgt, dann ergibt sich das *Gesetz von Clausius-Mosotti* als Zusammenhang zwischen Suszeptibilität und Polarisierbarkeit $\chi = 3\varkappa/(3 - \varkappa)$. Die Polarisierbarkeit $\varkappa = 3(\varepsilon - 1)/(\varepsilon + 2)$ kann daher bestimmt werden, wenn man die relative Dielektrizitätskonstante mißt.

Dielektrizität. So bezeichnet man die Polarisation, wenn die Dipolmomente erst durch das äußere Feld erzeugt werden. Es verschieben sich zum einen die Elektronenhüllen gegenüber den Atomkernen (*Elektronenpolarisation*), zum anderen verschieben sich die positiven Ionen gegenüber den negativen (*Ionenpolarisation*).

Parelektrizität liegt vor, wenn in der Substanz permanente Dipole vorhanden sind, bestehend aus polaren Molekülen oder aus Komplexen von elektrisch geladenen Gitterbausteinen. Ohne äußeres elektrisches Feld sind die Richtungen dieser Dipole durch die Wärmebewegung statistisch verteilt, das mittlere Dipolmoment je Volumenelement verschwindet. Bei Anlegen eines Feldes orientieren sich die Dipole an dessen Richtung. Die parelektrische Suszeptibilität ist im Unterschied zur dielektrischen temperaturabhängig, sie befolgt das *Curie-Weißsche Gesetz* $\chi = C/(T - T_c)$.

Ferroelektrizität. Bei ihr wird auch ohne äußeres elektrisches Feld eine permanente Polarisation beobachtet. Sie ist an bestimmte Kristallstrukturen gebunden, Beispiele sind Seignettesalz, Bariumtitanat, Triglyzinsulfat (TGS), Kaliumhydrogenphosphat (KDP). Die ferroelektrische Phase existiert nur in gewissen Temperaturbereichen, außerhalb dieser Umwandlungstemperaturen werden diese Substanzen parelektrisch. Die Suszeptibilität χ und damit die Dielektrizitätskonstante ε sind im ferroelektrischen Zustand stark feldabhängig, es tritt eine ferroelektrische Hysterese auf. Ferroelektrische Materialien werden z. B. für Kondensatoren großer Kapazität, als elektrooptische Wandler und als Strahlungsempfänger eingesetzt.

6.7.4. Piezoelektrizität

Eine Verformung in Richtung einer polaren Achse ruft bei piezoelektrischen Kristallen eine Polarisation und eine damit verbundene elektrische Spannung infolge von Ladungsverschiebungen hervor (*Piezoeffekt*). In Abb. 6.7.4–1 ist als Beispiel die Struktur von Quarz dargestellt. Im undeformierten Zustand sind die Sauerstoff- und Siliciumionen so zueinander angeordnet, daß der Kristall nach außen neutral ist. Bei einer Deformation mittels der mechanischen Spannung σ verschieben sich die Ionen wegen des fehlenden Inversionszentrums (vgl. 6.1.) so, daß die Probe polarisiert wird, wobei für die Polarisation $P = d\sigma$ gilt (d = materialabhängiger *Piezokoeffizient*). Wird umgekehrt ein elektrisches Feld E an den Kristall angelegt, so verformt sich dieser (*reziproker Piezoeffekt*). Für die Dehnung ε gilt $\varepsilon = dE$ mit demselben d wie für den Piezoeffekt. Während alle ferroelektrischen Kristalle (vgl. 6.7.3) auch piezoelektrisch sind, gilt das Umgekehrte nicht, wofür Quarz ein Beispiel ist. Angewendet werden piezoelektrische Substanzen als *elektromechanische Wandler* in vielfältigster Form, z. B. zur Messung von Kräften, als Tonabnehmer für Schallplatten (piezoelektrische Biegestreifen), als Mikrofonmembran (piezokeramische Dünnschicht auf Aluminiummembran), als Ultraschallstrahler, als Frequenznormal für Uhren, Sender und Frequenzmesser (Quarzkristalle werden durch Wechselspannung zu Eigenschwingungen angeregt; es wird eine Frequenzgenauigkeit $\Delta\nu/\nu = 10^{-8}$ erreicht!).

6.7.5. Pyroelektrizität

Kristalle, die piezoelektrisch sind und gleichzeitig ein permanentes elektrisches Dipolmoment bereits ohne Anwendung einer äußeren mechanischen Spannung längs der Kristallachse größter thermischer Ausdehnung aufweisen, zeigen Pyroelektrizität. Dieses Dipolmoment ist zwar i. allg. durch Oberflächenladungen kompensiert und nicht nachweisbar, die Ladungsverteilung an der Oberfläche ist aber so stabil, daß sie plötzlichen Änderungen der spontanen Polarisation nicht folgen kann. Eine solche Änderung bewirkt aber nun gerade eine homogene Temperaturänderung ΔT des Kristalles, wodurch die Polarisationsänderung ΔP meßbar wird, $\Delta P = p\Delta T$ (*Pyroelektrizität*). Der *pyroelektrische Koeffizient* p ist damit gleich dem Temperaturkoeffizienten der spontanen Polarisation des Materiales.

Beispiele für pyroelektrische Kristalle sind Turmalin, Bariumtitanat, Triglyzinsulfat (TGS). Sie werden als *Strahlungsdetektoren*, insbesondere im infraroten Spektralbereich, eingesetzt.

6.7.6. Komplexe frequenzabhängige dielektrische Funktion

Wirkt auf ein Elektron ein periodisches elektrisches Feld $\boldsymbol{E} = \boldsymbol{E}_0 e^{-i\omega t}$ (der Realteil $\boldsymbol{E}_0 \cos \omega t$ beschreibt das Feld), so wird es zu erzwungenen Schwingungen im Takt des äußeren Feldes angeregt (vgl. 2.1.5.3.), deren Amplitude und Phasenverschiebung gegen das erregende Feld frequenzabhängig sind. Das führt zu einer Polarisation $\boldsymbol{P} = en\boldsymbol{s} \equiv \chi(\omega)\varepsilon_0 \boldsymbol{E}$ bzw. zu einer elektrischen Stromdichte $\boldsymbol{j} = en\dot{\boldsymbol{s}} = \dot{\boldsymbol{P}} \equiv \sigma(\omega)\boldsymbol{E}$ (vgl. 6.7.1.), die durch eine frequenzabhängige Suszeptibilität $\chi(\omega)$ bzw. Leitfähigkeit $\sigma(\omega)$ beschrieben werden. In der komplexen Schreibweise kann die Phasenverschiebung leicht durch das Verhältnis von Real- zu Imaginärteil erfaßt werden, so daß man mit einer komplexen Suszeptibilität $\chi(\omega) = \chi_1(\omega) + i\chi_2(\omega)$ bzw. einer komplexen dielektrischen Funktion $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$ arbeitet. Ein mit der Frequenz ω_0 gebundenes Elektron liefert einen Beitrag $\chi \rightarrow (e^2/m\varepsilon_0)\ (\omega_0^2 - \omega^2 - i\omega/\tau)^{-1}$, wobei τ die mittlere Stoßzeit (vgl. 6.7.1.) darstellt.

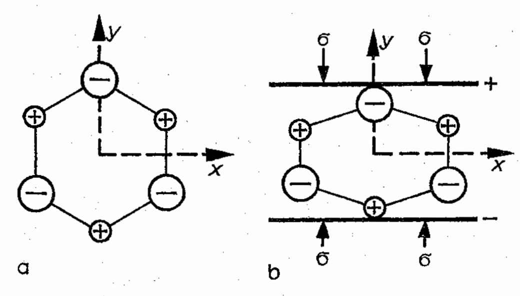


Abb. 6.7.4–1 Piezoeffekt; (a) Quarz in undeformiertem Zustand, (b) Polarisation durch Ladungsverschiebung infolge Deformation

Das elektrische Verhalten einer Substanz wird also durch zwei frequenzabhängige Funktionen beschrieben: 1. Real- und Imaginärteil der *Suszeptibilität* $\chi(\omega)$ oder 2. Real- und Imaginärteil der *dielektrischen Funktion* $\varepsilon(\omega) = 1 + \chi(\omega)$ oder 3. Real- und Imaginärteil des *komplexen Brechungsindex* $N(\omega)$, wobei die Definition $N^2(\omega) = \varepsilon(\omega)$ gilt, oder 4. *Brechungsindex* $n(\omega)$ *und Absorptionskoeffizient* $a(\omega)$, die aus $N(\omega) = n(\omega) + ia(\omega)$ hervorgehen. In analoger Weise ist eine komplexe magnetische Suszeptibilität einzuführen. Die beiden Funktionen χ_1 und χ_2 sind über die *Kramers-Kronig-Relationen* miteinander verknüpft, die aus der Kausalität des physikalischen Geschehens folgen. Nach Messung des Imaginärteiles $\chi_2(\omega)$ kann mit ihrer Hilfe $\chi_1(\omega)$ berechnet werden.

6.7.7. Magnetismus

Analog der Polarisation (vgl. 6.7.3.) induziert ein Magnetfeld in einer Substanz eine Magnetisierung $\boldsymbol{M} = \chi\boldsymbol{H}$. Die magnetische Suszeptibilität χ hängt entsprechend dem Clausius-Mossottischen Gesetz mit der Magnetisierbarkeit $\varkappa$ zusammen $\chi = 3\varkappa/(3 - \varkappa)$.

6.7.7.1. Diamagnetismus. In allen Atomen werden durch ein Magnetfeld atomare Ströme induziert, die einem magnetischen Moment entsprechen. Dieses ist dem äußeren Feld entgegengerichtet, d. h. die *diamagnetische Suszeptibilität* $\chi_{\text{dm}} = -N[\mu_0 Ze^2(6m)]\overline{r^2}$ ist negativ (N = Zahl der Atome je Volumen, Z = Ordnungszahl, m = Elektronenmasse, $\overline{r^2}$ = Mittelwert der Elektronenverteilung über die Elektronenwolke im Atom). Sie liegt betragsmäßig in der Größenordnung 10^{-5}. Diamagnetische Substanzen werden in einem inhomogenen Magnetfeld aus dem Bereich hoher Feldstärke herausgedrängt. Diamagnetische Kristalle sind vorwiegend Isolatoren, z. B. Steinsalz, Flourit, Calcit, Eis, aber auch Metalle, wie Kupfer und Bismut.

6.7.7.2. Paramagnetismus. Er wird beobachtet, wenn die Bausteine der Substanz ein permanentes magnetisches Moment aufweisen. Wie bei der Parelektrizität ist die Suszeptibilität temperaturabhängig und befolgt das Curie-Weißsche Gesetz $\chi = C/(T - T_c)$ (vgl. 4.10.1.). Während die freien Atome und Ionen i. allg. ein permanentes magnetisches Moment haben, wird bei deren Zusammenbau zum Festkörper die Elektronenstruktur so stark verändert, daß sich die magnetischen Momente der einzelnen Elektronen meist gegenseitig kompensieren. Die Ausnahmen bilden vor allem die Seltenen Erden, bei denen die nichtaufgefüllten f-Schalen soweit im Innern des Atoms liegen, daß sie durch die chemische Bindung kaum beeinflußt werden.

Die Leitungselektronen in Metallen tragen über ihr mit dem Elektronenspin verknüpftes Eigenmoment von der Größe eines Bohrschen Magnetons μ_B (vgl. 4.3.2.) und durch ihre Ladung zum Magnetismus bei. Beim Anlegen eines äußeren Magnetfeldes tritt ein Energieunterschied zwischen den Zuständen verschiedener Spinrichtung der Elektronen auf, was zu einer stärkeren Besetzung der Zustände führt, bei denen das magnetische Moment der Elektronen parallel zum Feld liegt. Dieser *Paulische Spinparamagnetismus* der Leitungselektronen führt zu einer Suszeptibilität $\chi = \mu_0\mu_B^2 D(E_F)$, worin $D(E_F)$ die Zustandsdichte an der Fermi-Energie (vgl. 4.10.2.) bedeutet. Der Elektronen-Paramagnetismus ist temperaturunabhängig. Die durch das Magnetfeld induzierte zusätzliche Bewegung der Elektronen führt zum *Landauschen Diamagnetismus*, für den in erster Näherung $\chi_{\text{Landau}} = -(1/3)\chi_{\text{Pauli}}$ gilt.

6.7.7.3. Ferromagnetismus. Unterhalb einer kritischen Temperatur, der *Curie-Temperatur* T_c, kann sich bei paramagnetischen Stoffen eine permanente Magnetisierung ohne äußeres Feld ausbilden. Die magnetischen Momente der Gitterbausteine oder nichtkompensierte Eigenmomente der Leitungselektronen (Bandmagnetismus) ordnen sich parallel. Aus energetischen Gründen erfolgt diese Parallelordnung nur in gewissen Bereichen, den *Weißschen Bezirken* bzw. *ferromagnetischen Domänen*. Diese sind im Mikroskop sichtbar (typische Ausdehnung etwa 10 µm). An den Grenzen der Weißschen Bezirke erfolgt der Übergang in eine andere Spinorientierung eines benachbarten Bezirkes nicht sprunghaft, sondern kontinuierlich in den *Bloch-Wänden* über eine Distanz von 100 bis 1 000 Gitterkonstanten. Im unmagnetischen Zustand kompensieren sich die magnetischen Momente der Weißschen Bezirke. Wird ein äußeres Magnetfeld angelegt, entsteht eine makroskopische Magnetisierung der Probe infolge zweier Mechanismen: Einmal wachsen diejenigen Weißschen Bezirke, deren Moment zum angelegten Magnetfeld günstig orientiert ist, durch Verschiebung der Bloch-Wände, zum anderen erfolgt eine Rotation der magnetischen Momente der Weißschen Bezirke in die Richtung des angelegten Magnetfeldes. Bei niedrigen Feldstärken dominiert der erste, bei hohen Feldstärken der letztere Mechanismus. Die Ausrichtungsvorgänge verlaufen unstetig und führen zu Sprüngen in der Magnetisierungskurve (*Barkhausen-Sprünge*).

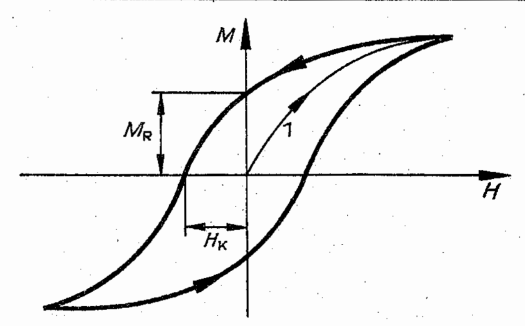


Abb. 6.7.7–1 Ferromagnetische Hysteresekurve

Der Zusammenhang zwischen Magnetisierung M der Probe und äußerem Magnetfeld H ist bei Ferromagnetika nicht mehr linear. Es tritt eine *ferromagnetische Hysterese* auf (Abb. 6.7.7–1). Bei Anlegen des Magnetfeldes H folgt die Magnetisierung M der jungfräulichen Kurve *1* bis zum Erreichen der *Sättigungsmagnetisierung*, bei der alle Weißschen Bezirke ausgerichtet sind. Wird das Magnetfeld bis $H = 0$ verringert, so bleibt eine gewisse Magnetisierung bestehen, die *Remanenz* M_R. Erst durch Anlegen des *Koerzitivfeldes* H_K entgegen der ursprünglichen Feldstärke wird der Wert Null der Magnetisierung erreicht. Werkstoffe mit großer Koerzitivfeldstärke H_K werden als magnetisch hart bezeichnet. Sie eignen sich besonders zur Herstellung von *Permanentmagneten*. Magnetisch weiche Werkstoffe (kleine Koerzitivfeldstärke H_K) haben eine schmale Hysteresekurve, d.h. die bei der Ummagnetisierung dissipierte Energie, die dem Flächeninhalt dieser Kurve entspricht, ist klein, so daß sie besonders als Werkstoff für Transformatorenbleche verwendet werden. Wichtige ferromagnetische Werkstoffe sind Eisen, Nickel, Cobalt sowie zahlreiche Legierungen, wie Permalloy, Alnico und insbesondere die Heuslerschen Legierungen.

6.8. Wechselwirkung mit Strahlung

Eine Grobeinteilung der Strahlung in *Strahlungsarten* ergibt elektromagnetische Strahlung (z. B. Licht, Röntgen-Strahlung, γ-Strahlung), Teilchenstrahlung geladener Teilchen (z. B. Elektronen, Ionen, Protonen) und Teilchenstrahlung ungeladener Teilchen (z. B. Neutronen). Die Wechselwirkung der Strahlung mit kondensierter Materie, also mit Atomkollektiven, unterscheidet sich von der Strahlungswechselwirkung mit isolierten Atomen. Dabei treten Effekte auf, die durch die dreidimensional periodische Anordnung der Atome in Kristallen (vgl. 6.1.) hervorgerufen werden.

6.8.1. Wechselwirkung mit elektromagnetischer Strahlung

Die Ausbreitung elektromagnetischer Wellen im Festkörper führt zu einer frequenzabhängigen *Strahlungsdämpfung*. Deshalb untersucht man die Wechselwirkung zwischen kondensierter Materie und elektromagnetischer Strahlung in einem breiten Frequenzintervall vom langwelligen Infrarot bis zu den harten γ-Strahlen (Tab. 6.8.1–1). Das Gebiet der Mikrowellen wird hier ausgeklammert (vgl. 6.9.).

6.8.1.1. Das elektromagnetische Wellenfeld im Festkörper. Die charakteristischen Größen des elektromagnetischen Wellenfeldes sind die elektrische Feldstärke $\boldsymbol{E}$ und die magnetische Feldstärke $\boldsymbol{H}$ (vgl. 2.4., 6.7.). Die Feldvektoren werden von kondensierter Materie materialspezifisch beeinflußt. Wichtige Materialparameter sind die elektrische Leitfähigkeit γ (Metalle, Halbleiter, Isolatoren), der Magnetismus und die Realstruktur.

Die Wechselwirkung zwischen Strahlung und kondensierter Materie kann mit verschiedenen theoretischen Modellen beschrieben werden. Die Grundlage bilden die *Kontinuumstheorie* und die atomistische Theorie. Die mathematische Grundlage beider Theorien sind die *Maxwellschen Gleichungen* und die mit ihnen verknüpften Materialgleichungen (vgl. 2.4., 6.7.). Aus der Lösung der Maxwellschen Gleichungen folgt, daß sich die Komponenten der Feldvektoren $\boldsymbol{E} = (E_x, E_y, E_z)$ und $\boldsymbol{H} = (H_x, H_y, H_z)$ wellenförmig ausbreiten. Die einfachste Form einer „Wellensonde" ist eine homogene ebene Welle. Für diese sind die Feldgrößen nur von einer Ortskoordinate abhängig. Man kann sie in guter Näherung realisieren, indem man die von einem Hertzschen Dipol (z. B. angeregte Atome oder Moleküle) emittierten elektromagnetischen Kugelwellen in einem Abstand r betrachtet, der groß ist gegen die Wellenlänge der Strahlung.

Die Ausbreitung einer ebenen Welle wird durch den komplexen, frequenzabhängigen Brechungsindex $N = n + i\varkappa$ bestimmt (vgl. 6.7.6.). Ihre Phasengeschwindigkeit ist $v = c/n$, wenn c die Vakuumlichtgeschwindigkeit ist. Der Absorptionsindex $\varkappa$ bestimmt die Dämpfung der Welle. Nach Passieren eines Materials der Dicke d ist die Intensität der elektromagnetischen Strahlung um einen Faktor $\exp(-4\pi\varkappa d/\lambda_0) = \exp(-Kd)$ vermindert worden. Darin bedeutet λ_0 die Vakuumwellenlänge. K wird als Absorptionskoeffizient bezeichnet. Für $\varkappa \neq 0$ ergibt sich auch eine Phasenverschiebung α zwischen den Feldvektoren $\boldsymbol{E}$ und $\boldsymbol{H}$, $\tan\alpha = \varkappa/n$. Für Leiter wird der Absorptionsindex bei nicht zu hohen Frequenzen – aber bis in den Lichtwellenbereich hinein – im wesentlichen durch die Leitfähigkeit γ bestimmt $\varkappa = \mu_r\gamma/fn$.

6.8.1.2. Einfluß der Anisotropie. Eine charakteristische Eigenschaft kristalliner Festkörper ist ihre Anisotropie (vgl. 6.1.). Für Kristalle nichtkubischer Symmetrie wird aus der skalaren Größe ε_r ein Tensor zweiter Stufe (vgl. 13.1.1., 6.7.) $\{\varepsilon_{ik}\}$. Mit den Maxwellschen Gleichungen und der Anisotropie des dielektrischen Verhaltens sind die physikalischen Grundlagen der Wechselwirkung zwischen elektromagnetischer Strahlung und kondensierter Materie im Rahmen der Kontinuumstheorie vollständig gegeben. Damit sind nicht alle Phänomene erfaßt, wie z. B. die elastische Lichtstreuung (Rayleigh-Streuung, vgl. 6.8.1.4.) und die Undurchsichtigkeit von Isolatoren.

6.8.1.3. Einfluß der atomistischen Struktur. Die frequenzabhängige Streuung elektromagnetischer Wellen durch feste Körper ist ein deutlicher Hinweis darauf, daß die Strahlung die atomistische Struktur spürt. Die Dispersion hat ihre Ursache in der Vergleichbarkeit der Wellenlängen der Strahlung mit dem Durchmesser der Atome im Festkörper sowie der Frequenzen der Wellen mit den Eigenfrequenzen der schwingungsfähigen Bausteine der kondensierten Materie. Jedes Atom bzw. Molekül bildet einen elektrischen und einen magnetischen Dipol mit den Dipolmomenten $\boldsymbol{p}$ bzw. $\boldsymbol{m}$. Es gilt $\boldsymbol{p}_i = \sum q_i \boldsymbol{r}_i = q\boldsymbol{R}$ (vgl. 6.7.). Ein äußeres Feld führt zu einer Ausrichtung bereits vorhandener Dipole in Feldrichtung und damit zu einer *Polarisation* $\boldsymbol{P}$ bzw. *Ma-*

Tab. 6.8.1-1 Das elektromagnetische Spektrum (ohne Radiowellen und Wechselströme)

Strahlung	Wellenlänge in cm	Frequenz in s^{-1}	Energie in eV
γ (Gamma)	$10^{-11}-10^{-8}$	$3\cdot10^{21}-3\cdot10^{18}$	$1{,}4\cdot10^{6}-1{,}4\cdot10^{3}$
X (Röntgen)	$5\cdot10^{-10}-10^{-5}$	$6\cdot10^{19}-3\cdot10^{15}$	$2{,}8\cdot10^{5}-13{,}9$
UV	$10^{-6}-4\cdot10^{-5}$	$3\cdot10^{16}-7{,}5\cdot10^{14}$	$1{,}4\cdot10^{2}-3{,}5$
Licht (sichtbar)	$4\cdot10^{-5}-8\cdot10^{-5}$	$7{,}5\cdot10^{14}-3{,}8\cdot10^{14}$	$3{,}5-1{,}7$
UR (IR)	$8\cdot10^{-5}-10^{-1}$	$3{,}8\cdot10^{14}-3\cdot10^{11}$	$1{,}7-1{,}4\cdot10^{-3}$
Mikrowellen	$10^{-2}-1$	$3\cdot10^{12}-3\cdot10^{10}$	$1{,}4\cdot10^{-2}-1{,}4\cdot10^{-4}$

Spektroskopische Methoden – anregende Strahlung

γ	*X*	*UV*	*Licht*	*UR*	*Mikrowellen*
Compton-Spektroskopie (a) Paarbildung (b) Mössbauer-Spektroskopie (c)	Emission/Absorption Röntgen-Strukturanalyse (a) Diffraktometrie, Kossel-Interferenz (b) Emissionsspektroskopie (c) Absorptionsspektroskopie (d) Extended X-ray-Absorption Fine Structure (EXAFS) (e) Compton-Spektroskopie (f) Röntgen-Fotoelektronenspektroskopie (XPS) (g) Auger-Elektronenspektroskopie (AES) (h) Röntgen-Fluoreszenzanalyse (RFA) (i)	Reflexion/Absorption, Elektronenschwingungsspektroskopie (a) Ultraviolett-Fotoemissions-Spektroskopie (UPS) (b) UPS/Synchrotronstrahlung (c)	Reflexion/Absorption, Raman-Spektroskopie, Magnetoabsorption, Faraday-Spektroskopie, Rayleigh-Spektroskopie, Lumineszenz (a) Fotoelektronenspektroskopie (b) Fotoleitung (c) Ellipsometrie (d)	Absorption/Reflexion (a) Raman-Spektroskopie (b) Zyklotronresonanz (c) Magnetoabsorption (d) Faraday-Spektroskopie (e) Rayleigh-Spektroskopie (f)	Electron Paramagnetic Resonance (paramagnetische Elektronenresonanz) (EPR) (a) Nuclear Magnetic Resonance (Kernresonanz) (NMR) (b) Zyklotronresonanz (c)

Informationen über Festkörpereigenschaften

Impulsverteilung der Valenzbandelektronen (Wellenfunktionen) (a) Positronenannihilation (Elektronenimpulsdichte-Verteilung, Leerstellenkonzentration, Leerstellenbildungsenthalpie, Versetzungsdichte (b) Ladungsdichteverteilung, chemische Wertigkeit, Gittersymmetrieverzerrungen, Phasenumwandlungen, magnetische Gitterplätze (c)	innere und äußere Elektronenübergänge, Atomabstände im Kristall, Kristallsymmetrie, Elektronendichte-Verteilung, Gitterschwingungen (a) Versetzungen, Stapelfehler (b) besetzte Zustände des Valenzbandes (c) unbesetzte VB-Zustände (d) Nahordnung (e) Elektronenimpulsverteilung, Plasmonen, Exzitonen (f) Elektronenstruktur, chemische Bindung, Oberflächenanalyse (g, h) Analyse (i)	äußere Elektronenübergänge, Vibronics (Gitterschwingungen) (a) Valenzelektronenbänder, Oberflächenzustände, Oberflächenanalyse (b) Oberflächenreaktionen (b, c) Elektronenladungsverteilung (c)	Elektronenstruktur äußerer Energieniveaus, Donator- und Akzeptorzentren (a) Elektronenaustrittsarbeit (b) Elektronenstruktur (c) optische Konstanten (Polarisation, Relaxation, Absorption, freie Elektronen, Plasmonen, Interbandübergänge (d)	Gitterschwingungen (a) innere Gitterschwingungen in Molekülkristallen, Molekülrotationen, Adsorbatstrukturen (b) IR-inaktive Schwingungen (c) Ladungsträgerkonzentration, Dielektrizitätskonstante (d, e) Kristallbaufehler (f)	Elektronenspinübergänge (ungepaarte Elektronen), Übergangsmetallverbindungen, offene Bindungen (freie Radikale) (a) Kernspinübergänge (Wechselwirkung benachbarter Protonen), Detailstruktur gestörter Einkristalle (b) effektive Masse, Relaxationszeit von Elektronen (c)

gnetisierung M. Für einen aus N Atomen je Volumen ΔV aufgebauten Festkörper ist $P = (1/\Delta V) \sum_{i=1}^{N} p_i$ und $M = (1/\Delta V) \sum_{i=1}^{N} m_i$. Die temperaturbedingten Schwankungen von P und M um einen Mittelwert bewirken die Streuung elektromagnetischer Wellen.

Wenn keine permanenten atomaren Dipole vorhanden sind, reagiert der Festkörper in einem elektromagnetischen Strahlungsfeld mit der Bildung atomarer Momente. Polarisation und Magnetisierung sind linear abhängig von den Erregerfeldstärken E und H. Es gilt $P = \chi_e E$ und $M = \chi_m H$. χ_e ist die *elektrische*, χ_m die *magnetische Suszeptibilität*.

Die Größe des *atomaren Dipolmomentes* p_i ist abhängig von der Feldstärke des lokalen elektrischen Feldes am Ort des i-ten Atoms: $p_i = \alpha\varepsilon_0 E_i$. Der Faktor α ist die *Polarisierbarkeit* des Atoms. Sie setzt sich zusammen aus einem elektronischen, einem ionischen und einem permanenten Dipolanteil.

Die verschiedenen Beiträge zur Polarisierbarkeit haben eine unterschiedliche Frequenzabhängigkeit. Die Umorientierung der permanenten Dipole erfolgt bis zu Frequenzen des Mikrowellenbereiches (vgl. 6.9.). Für höhere Frequenzen (z. B. IR) sind die Dipole zu träge und werden von dieser „Strahlensonde" nicht mehr bemerkt.

Die Ionenpolarisation endet im Infrarotbereich, während die Elektronenpolarisation von allen Frequenzen gespürt wird.

6.8.1.4. Wechselwirkung mit Infrarot-Strahlung. Absorptionsspektrum. Mit Hilfe des Spektrums der elektromagnetischen Strahlung (vgl. Tab. 6.8.1–1) und des Absorptionskoeffizienten K läßt sich ein Diagramm der Absorptionseigenschaften des Festkörpers angeben. Der einfachste Fall ist ein Isolator ($\gamma \simeq 0$) mit idealem Gitteraufbau. Charakteristisch ist die Zweiteilung des Spektrums (Abb. 6.8.1–2). Die Strahlung großer Wellenlängen regt im Festkörper Gitterschwingungen (vgl. 6.4.) an. Die Strahlungsenergie wird in Wärmeenergie des Festkörpers umgewandelt. Am Schmelzpunkt des Festkörpers wird das Kristallgitter zerstört, d. h. die Fernordnung der Atome bzw. Ionen ist nicht mehr vorhanden. Der kurzwellige Teil des Absorptionsspektrums ist deutlich strukturiert. Er entsteht durch strahlungsinduzierte Elektronenübergänge zwischen atomaren Energieniveaus oder breiten Energiebändern der Valenzelektronen (Interbandübergänge). Dieser zweigeteilte Absorptionsmechanismus bildet die *Grundgitterabsorption*. Wegen der Absorptionslücke im Spektralbereich des sichtbaren Lichtes (vgl. Abb. 6.8.1–1 und –2) müßte jeder Isolator durchsichtig sein. Das ist nicht richtig. Deshalb muß es zumindest einen weiteren Mechanismus geben, der durch die Abweichung des Realkristalles vom Idealkristall entsteht. Dieser als *Ausläuferabsorption* bezeichnete Mechanismus ist ebenfalls ein elektronischer, allerdings verursacht durch Elektronen von Gitterstörungen, wie Verunreinigungsatomen, Gitteratomen auf nicht regulären Plätzen und Elektronen auf unbesetzten Gitterplätzen. Die Ausläuferabsorption ($K_{max} \simeq 1\,mm^{-1}$) schließt die Absorptionslücke zwischen den beiden Anteilen zur Grundgitterabsorption überlappend. *Intrabandabsorption*. Bei Metallen und Halbleitern im entarteten Zustand (verbotene Zone $E_g \lesssim k_B T$) gibt es weiterhin die Möglichkeit der Absorption durch freie Elektronen im Leitungsband. Innerhalb dieses Bandes werden nach der Absorption höherenergetische Zustände besetzt. Für die Wellenlängenabhängigkeit der Absorption durch freie Ladungsträger gilt: $K \sim \lambda^m (1{,}5 \leqq m \leqq 3{,}5)$, woraus ein hoher IR-Anteil folgt. Der Wertebereich für m ist ein Hinweis auf verschiedene Wechselwirkungsmechanismen der angeregten Elektronen mit dem Kristallgitter. Die kleinen Werte sind charakteristisch für die Beteiligung von Phononen, die großen für die Wechselwirkung mit geladenen Störstellen.

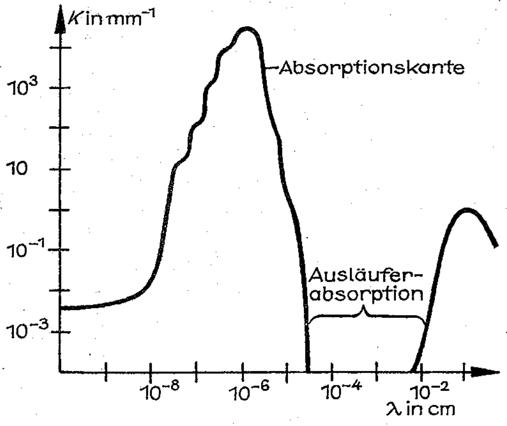


Abb. 6.8.1–2 Absorptionsspektrum des idealen Festkörpers

Ultrarot-Spektrometrie. Sie wird als Absorptionsspektrometrie (URA) und als Reflexionsspektrometrie (URR) durchgeführt. Das UR-Spektrum umfaßt den Wellenlängenbereich $6 \cdot 10^6 \leqq \lambda/\mathrm{nm} \leqq 6 \cdot 10^3$. Das entspricht Wellenzahlbeträgen $|q| = 2\pi/\lambda$ von $10^{-6} \leqq |q|/\mathrm{rad}^{-1}\mathrm{nm} \leqq 10^{-4}$, Kreisfrequenzen $\omega = 2\pi f$ von $3 \cdot 10^{12} \leqq \omega/\mathrm{rad}_s^{-1} \leqq 3 \cdot 10^{14}$ und Energien $E = \hbar\omega$ von $2 \cdot 10^{-3} \leqq E/\mathrm{eV} \leqq 2 \cdot 10^{-1}$.

Diese Strahlungsparameter bedingen eine starke Wechselwirkung der UR-Strahlung mit den *Gitterschwingungen* (vgl. 6.4.). Im Teilchenbild sind die Wechselwirkungen unelastische und elastische Stoßprozesse zwischen Photonen und Phononen. Die Photonen (bzw. elektromagnetische Wellen) können nur in Kristallgittern mit elektrischen Dipolen Gitterschwingungen anregen. Die Absorption erfolgt durch elektrische Dipole. Das ist besonders bei Ionenkristallen und Halbleitern mit hohem Ionenbindungsanteil möglich. Je nach dem Aufbau der Kristalle können den angeregten eigentlichen Gitterschwingungen, wie sie bei Gittern aus nichtmolekularen Bausteinen (z. B. Na^+, Cl^-) vorkommen, *innere Gitterschwingungen* überlagert sein. Diese Erscheinung tritt bei Molekülkristallen (Sulfate, Carbonate, Nitrate u. a.) auf. In den Komplexen SO_4^{--}, CO_3^{--}, NO_3^- werden durch die UR-Strahlung innere Schwingungen der Bestandteile der Komplexe gegeneinander angeregt. Dabei schwingt der Komplex unter Umständen auch noch als Ganzes gegen die übrigen Bestandteile des Gitters synchron. Bei gegenphasigem Schwingungsverhalten von entgegengesetzt geladenen Gitterbausteinen (z. B. Na^+Cl^-) ist die Änderung der Dipolmomente besonders groß, die entsprechende Schwingung besonders ultrarotaktiv. Eine direkte Anregung von Gitterschwingungen durch UR-Photonen ist nur in optischen Zweigen möglich. Indirekt besteht die Anregungsmöglichkeit für akustische Phononen durch eine nichtlineare Kopplung (d. h. außerhalb des Gültigkeitsbereichs des Hookeschen Gesetzes) mit optischen Phononen. Im Absorptionsspektrum eines Festkörpers tritt eine Vielzahl von breiten Absorptionsmaxima, *Absorptionsbanden*, auf (Abb. 6.8.1–3). Die Absorptionsbanden sind Kombinationen einer geringeren Zahl von *Fundamentalschwingungen*.

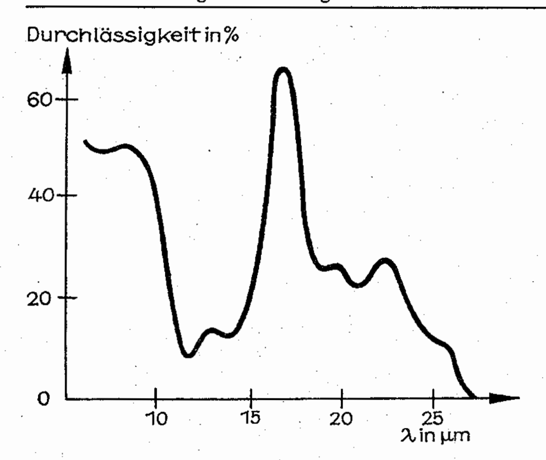


Abb. 6.8.1–3 Absorptionsspektrum von Galliumarsenid GaAs im UR-Spektralbereich

Die Fundamentalschwingungen gehören zu einem optischen (O) Zweig oder zu einem akustischen (A) Zweig und können transversal (T) oder longitudinal (L) polarisiert sein (vgl. 6.4.). Folgende Kombinationsschwingungen (darunter auch Oberschwingungen) können z. B. direkt im Absorptionsspektrum gemessen werden: TO; (LO + LO); (TO + TO); (LO + LA); (LO + TA_i) (*i*: Zweigindex 1; 2; ... usw.). Bei Berücksichtigung der Symmetrieelemente und der Symmetriegruppen der Kristalle gelingt es mit gruppentheoretischen Berechnungen (vgl. 6.4.) die Wellenzahlen der Fundamentalschwingungen zu berechnen. Der Vergleich mit den gemessenen Absorptionsbanden führt zu den Kombinationsschwingungen. Auch bei streng kovalenten Gittern (Germanium, Silicium) können ultrarotaktive Schwingungen durch Symmetrieverzerrungen im Kristallgitter beobachtet werden. Die optische Aktivität ist jedoch wesentlich kleiner als bei hohen ionischen Bindungsanteilen.

Im kurzwelligen UR und bei schmalbandigen Halbleitern werden durch Strahlungswechselwirkung Elektronen vom Valenzband in das Leitungsband (vgl. 6.3.1.) angehoben. Da der Impuls des anregenden Photons praktisch Null ist, können nur Zustände im Leitungsband besetzt werden, die den gleichen Elektronenimpuls haben wie der Ausgangszustand im Valenzband. Diese *direkten Übergänge* sind ziemlich unwahrscheinlich wegen der durch den Impulserhaltungssatz gegebenen Einschränkung. Durch die Beteiligung von Gitterschwingungen am Energietransfer werden wegen der wesentlich größeren „Impulse" der Phononen auch andere Zustände im Leitungsband besetzt. Dabei kann die Photonenabsorption (möglicher direkter Übergang ΔE) sowohl von einer Phononenemission (tatsächlicher Energiesprung $\Delta E - E_{\text{Phonon}}$) als auch von einer Phononenabsorption (Energiesprung $\Delta E + E_{\text{Phonon}}$) begleitet sein. Diese *indirekten Übergänge* sind experimentell nachweisbar durch die Messung der Frequenzabhängigkeit des Absorptionskoeffizienten $K = f(\omega)$. Die Funktion $\omega = f(\sqrt{K})$ ergibt sich bei indirekten Übergängen aus einer Superposition von zwei Geraden, deren Achsenabschnitte auf der Ordinate die Werte für ΔE und E_{Phonon} liefern.

Einen weiteren Beitrag zum Absorptionsspektrum liefert die Anregung von Elektronen-Loch Paaren (*Excitonen*), deren Coulombsche Wechselwirkungskraft sich aus einem wasserstoffähnlichen Energietermschema der Form $E_n = E_\infty(1 - 1/n^2)$ ergibt. Das spiegelt sich im Spektrum wider (Abb. 6.8.1–4).

Um die Überlagerung dieses Spektrums mit dem der Gitterschwingungen zu umgehen, wird es bei sehr tiefen Temperaturen (~4 K) aufgenommen.

Die *Zyklotronresonanz* (vgl. 6.9.) läßt sich auch mit ultraroter Strahlung nachweisen. Durch eine Registrierung der magnetfeldaufgespalteten *Plasmaresonanzkante* (vgl. 6.8.1.6.) ist die Ladungsträgerkonzentration und die Dielektrizitätskonstante des Festkörpers bestimmbar. Die Wechselwirkung elektromagnetischer Strahlung mit kondensierter Materie für Frequenzen, die nicht absorbiert werden, aber in der Nähe der Absorptionskante liegen, kann zum Nachweis des *Zeeman-Effektes* im Festkörper benutzt werden. Die durchgelassene Strahlungsintensität ist durch *Magnetoabsorptionen* moduliert. Die Intensitätsoszillationen sind von der Lage des angelegten Magnetfeldes abhängig. Die Oszillationen entstehen durch die Aufspaltung der elektronischen Energieniveaus im Magnetfeld, den sog. *Landau-Niveaus*. Die Bestimmung von effektiven Massen und Ladungsträgerkonzentrationen im Festkörper gelingt auch durch den *Faraday-Effekt*. Die Drehung der Polarisationsebene linear polarisierten Lichtes unter Einfluß eines äußeren Magnetfeldes kann sowohl in Transmission als auch in Reflexion gemessen werden. Weitere Wechselwirkungsphänomene zwischen elektromagnetischer Strahlung und dem Festkörper werden durch die *Rayleigh-Streuung* (*Tyndal-Effekt*), die *Raman-Streuung* und die *Brillouin-Streuung* erfaßt. Die *Rayleigh-Streuung* ist eine Zerstreuung des Lichtes durch Kristallbaufehler, deren molekulare Struktur klein ist gegen die Lichtwellenlänge im Ultrarotbereich. Die lokale Änderung der Polarisierbarkeit bewirkt einen elastischen Streuprozeß, die Wellenlänge von einfallendem und gestreutem Licht ist identisch. Die gestreute Intensität ist wellenlängenabhängig: $I_{\text{streu}} \sim 1/\lambda^4$. Bei bestimmten Kristallsymmetrien kann die Fundamentalschwingung, die eine maximale Änderung des resultierenden Dipolmomentes im Festkörper hervorruft ($\Omega^{\text{TO}}(q = 0)$), optisch inaktiv sein (z. B. bei Calciumfluorid). Mit Hilfe des Raman-Effektes gelingt dennoch ein experimenteller Nachweis dieser Schwingung. Der *Raman-Effekt* beruht als optisch aktiver Effekt auf der zeitlichen Änderung der Polarisierbarkeit der Gitterbausteine, speziell der Elektronenhüllen. Sie entsteht durch die Änderung der Schwingungsamplitude der Gitterschwingungen, bzw. im Teilchenbild durch die Erzeugung

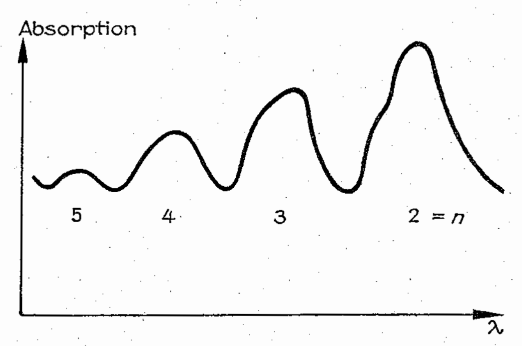


Abb. 6.8.1–4 Beitrag zur Absorption von UR-Photonen durch Excitonenanregung

oder Vernichtung von Phononen. Der Streuvorgang variiert die eingestrahlte Frequenz ω_0 der elektromagnetischen Welle gemäß $\omega' = \omega_0 \pm \Omega^{(i)}(q)$. Dabei bedeuten $\Omega^{(i)}$ die Phononenfrequenzen der Zweige (i) und q die Phononenwellenzahl. Phononen des optischen Zweiges verursachen die Raman-Streuung, die der akustischen Zweige die Brillouin-Streuung. Beim Raman-Effekt 1. Ordnung geschieht die Wechselwirkung zwischen Photonen und Phononen mit Grundschwingungen, beim Raman-Effekt 2. Ordnung mit Kombinations- und Oberschwingungen. Die Raman-Spektroskopie mit hohem räumlichen Auflösungsvermögen gestattet den Nachweis von Atom- bzw. Molekülgruppen geringer Dimension entsprechend der Wellenlänge des eingestrahlten Lichtes.

Reflexionsspektrum. Die durch elektromagnetische Strahlung angeregten Dipolschwingungen im Festkörper können wie Antennen ihre Energie in Form elektromagnetischer Strahlung wieder abgeben (Wärmestrahlung). Dem Maximum der Absorption im Frequenzspektrum (vgl. 6.8.1–3) entspricht ein Maximum des Reflexionsvermögens. Letzteres ist gegenüber dem Absorptionspeak nach höheren Wellenlängen verschoben (Abb. 6.8.1–5).

Das liegt daran, daß das Reflexionsvermögen außer durch den Absorptionskoeffizienten K auch durch den Brechungsindex n (vgl. 6.8.1.) beeinflußt wird. Demnach ist die Ursache für die Verschiebung zwischen den korrelierten

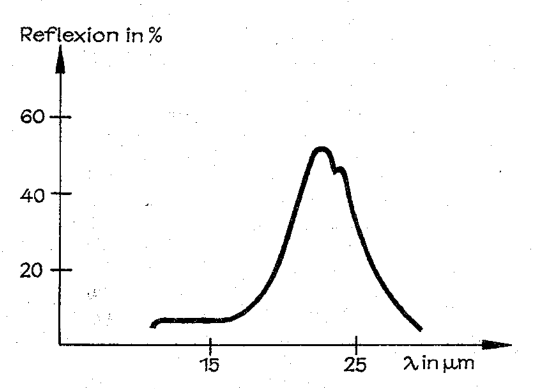


Abb. 6.8.1–5 Reflexionsspektrum von Galliumarsenid GaAs im UR-Bereich

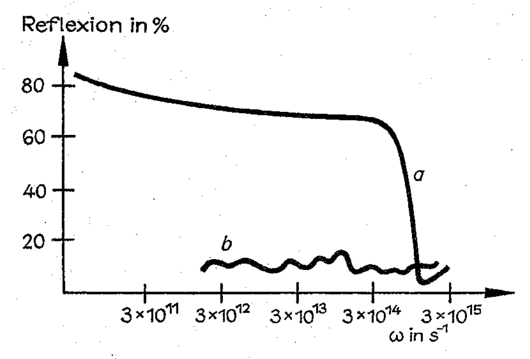


Abb. 6.8.1–6 Reflexionsvermögen im UR-Bereich von (a) Metallen und (b) Isolatoren

Maxima von Absorption und Reflexion eine starke Wechselwirkung zwischen der anregenden Welle und der angeregten Gitterwelle im Resonanzfall. Analog zum Absorptionsverhalten beeinflussen die freien Elektronen des Festkörpers auch sein Reflexionsverhalten entscheidend. Die Reflexionsspektren von Isolatoren und Metallen unterscheiden sich erheblich voneinander (Abb. 6.8.1–6).

Die dielektrische Funktion $\varepsilon(\omega) = 1 - (\omega_p/\omega)^2$ führt für $\varepsilon < 0$ zur totalen Reflexion elektromagnetischer Wellen. Die Plasmafrequenz ω_p ist abhängig von der Konzentration der freien Ladungsträger n und deren effektiver Masse m_{eff}: $\omega_p = ne^2/\varepsilon_0 m_{\text{eff}}$. An der Plasmaabsorptionskante können Kollektivschwingungen des Ensembles der freien Elektronen gegen den „Untergrund" des aus positiven Ionen bestehenden Gittergerüstes des Festkörpers angeregt werden. Die elektromagnetische Welle wird dann vom Festkörper „angenommen".

6.8.1.5. Wechselwirkung mit sichtbarem Licht. Das sichtbare Licht (und der langwellige Teil des UV-Spektrums) fallen gerade in die Lücke der Ausläuferabsorption (vgl. 6.8.1.4. und Abb. 6.8.1–2). Also werden besonders Wechselwirkungseffekte wirksam, die durch Gitterstörungen (für Festkörper mit geringer elektrischer Leitfähigkeit) und durch freie Ladungsträger (bei hoher elektrischer Leitfähigkeit) hervorgerufen werden. Wie bei der UR-Strahlung ist auch für Licht (und für UV-Strahlung) die Wellenlänge der Erregerstrahlung groß gegen die Atomabstände im Festkörper ($\approx 10^{-8}$ cm). Die Raman- bzw. Brillouin-Streuung kann ebenso wie im Ultrarot mit sichtbarer oder UV-Strahlung beobachtet werden. Das gleiche gilt für die Rayleigh-Streuung, den Zeeman-Effekt und die Faraday-Rotation. Die Phänomene der Wechselwirkung von Licht und Materie werden in der Technik viel genutzt. Wegen der diskontinuierlichen Struktur der Festkörper sind sie immer eine Wechselwirkung zwischen Atomen und Photonen. Ein Strahlungsfeld, das aus n_A Photonen besteht, kann ein Photon an ein Atom „abgeben", wodurch letzteres in einen angeregten Energiezustand E_a übergeht (Photonenabsorption). Dieser gegenüber dem Grundzustand E_g höherenergetische Zustand hat eine bestimmte Lebensdauer τ. Sie beträgt im sichtbaren Spektralbereich für einen erlaubten elektrischen Dipolübergang $\tau \sim 10^{-8}$ s. Mit der Wahrscheinlichkeit $W = 1/\tau$ kann das Atom ohne äußere Einwirkung vom Zustand E_a in den Zustand E_g zurückkehren. Dabei wird ein Photon der Energie $\hbar\omega = E_a - E_g$ emittiert. Dieser Vorgang ist die *spontane Emission*. Eine andere Form der Emission ist die *induzierte Emission*, die in der Laser-Technik angewendet wird (vgl. 4.6.).

Lumineszenz ist eine spontane, kalte Lichtemission (keine Wärmestrahlung) als Folge vorangegangener Bestrahlung eines Festkörpers. Im Unterschied zur einfachen Reflexion, bei der die Wellenlänge und die Phasenlage des Lichts unverändert bleiben, gibt es bei der Lumineszenz-Remission Änderungen zwischen einfallender und remittierter Strahlung. Wegen der Frequenzverschiebung kann man den Effekt zum Nachweis verschiedenster Primärstrahlungen nach einem Meßprinzip (Sekundärelektronenvervielfacher) ausnutzen. Wie die einfache Reflexion basiert auch die Lumineszenz auf dem Elektronenübergang von angeregten Zuständen in energetisch niedere Zustände. Beim Lumineszenzvorgang sind aber stets Störstellen an einem Elektronenübergang beteiligt. Ein störstellenfreier Kristall zeigt keine Lumineszenz. Der Begriff Lumineszenz vereinigt zwei Erscheinungen in sich, die der *Fluoreszenz* und der *Phosphoreszenz*. Beide unterscheiden sich in der Dauer der

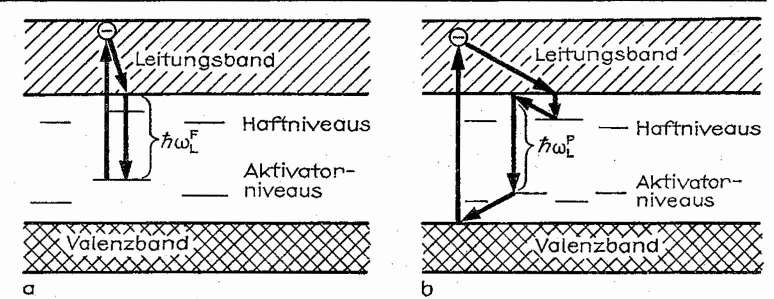


Abb. 6.8.1–7 Energiediagramme zur Entstehung der (a) Fluoreszenz- und (b) Phosphoreszenzstrahlung

Lichtemission nach der Wechselwirkung mit der Anregungsstrahlung. Während die Fluoreszenz nur ein Nachleuchten unmittelbar nach der Bestrahlung beschreibt, gibt es bei der Phosphoreszenz ein Nachleuchten von mindestens 10^{-3} s.

Verantwortlich für das unterschiedliche Zeitverhalten ist der Entstehungsmechanismus der Strahlung (Abb. 6.8.1–7). Daran beteiligt sind neben dem Valenz- und dem Leitungsband noch verschiedene Störstellen, die *Aktivatorniveaus* (Leuchtzentren) und *Haftniveaus* bilden. Letztere befinden sich energetisch in der Nähe der unteren Kante des Leitungsbandes.

Die Störstellenniveaus entstehen durch das gezielte Einbringen (Dotieren) von Fremdatomen oder durch die Erzeugung einer bestimmten Eigenfehlordnung im Grundgitter (Matrix) des Kristalles. Durch die Lichtabsorption gelangt ein Elektron eines Aktivatoratoms in einen unbesetzten Zustand des Leitungsbandes. Beim Absorptionsprozeß erfolgt die Umwandlung von Strahlungsenergie in Elektronenenergie. Das Elektron entfernt sich von „seinem" Atom. Durch Wechselwirkung mit dem Kristallgitter verliert es einen Teil seiner kinetischen Energie: Im Energiediagramm sinkt das Elektron an die untere Kante des Leitungsbandes (Abb. 6.8.1–7a). Von dort aus kommt es unter Aussendung von Fluoreszenzstrahlung (Frequenz ω_L^F) zur Rekombination des Elektrons mit dem Lochzustand des Aktivatoratoms. Der Vorgang der Fluoreszenz läuft in einem Zeitraum $<10^{-3}$ s ab. Bei der Phosphoreszenz ist der Rekombinationsmechanismus komplizierter (Abb. 6.8.1–7b). Nach der Strahlungsanregung von Elektronen aus dem Valenzband und der Wechselwirkung der nahezu freien Elektronen mit dem Kristallgitter (Absinken an die untere Kante des Leitungsbandes) erfolgt nicht sofort eine Rekombination an einem Aktivatorniveau. Das im Gitter diffundierende Elektron „sucht" sich erst ein freies Niveau, wobei es zwischenzeitlich von Haftstellen eingefangen wird. Je nach der Temperatur des Kristalles verweilt es mehr oder weniger lange dort. Bei niedrigen Temperaturen wird die Wahrscheinlichkeit der Strahlungsrekombination wegen der behinderten Diffusion kleiner sein. Daraus ergibt sich ein längeres Nachleuchten bei der *Tieftemperaturphosphoreszenz* im Vergleich mit der bei höheren Temperaturen. Zur *Phosphoreszenzanregung* nimmt man die Strahlung des langwelligen Ausläufers der Grundgitterabsorption. Die Grundgitterabsorption verursacht bei kürzeren Wellenlängen in einer dünnen Oberflächenschicht des Festkörpers hohe Elektronen- und Löcherdichten und somit eine hohe Rekombinationswahrscheinlichkeit zugunsten der einfachen Reflexion. Als Kristallphosphore sind am bekanntesten Zinksulfid (ZnS) und Cadmiumsulfid (CdS). Die Aktivatoratome sind Kupfer (Cu), Mangan (Mn), Silber (Ag), aber auch ein Überschuß von Zn-Atomen in ZnS (Nichtstöchiometrie) bildet Leuchtzentren. Die Aktivatoratome können dabei reguläre Gitterplätze aber auch Zwischengitterplätze einnehmen, wenn es die Größenverhältnisse der Atome zulassen. Die optimale Aktivatorkonzentration für eine intensitätsreiche Phosphoreszenzstrahlung beträgt bei Zwischengitterplatzbesetzung $\approx 10^{-5}$ (1 Leuchtzentrum auf 10^5 Gitteratome), bei Substitution $\approx 10^{-2}$. Technische Anwendungen der Lumineszenz sind Leuchtstofflampen, Leuchtfarben (Waschmittelindustrie), TV-Bildröhren, Lumineszenzdioden, Bildwandler, Meßgeräte der Kernstrahlungsmeßtechnik.

Äußerer lichtelektrischer Effekt. Bestrahlt man Festkörperoberflächen mit sichtbarem Licht (oder mit energiereicherer UV- oder Röntgen-Strahlung), so werden aus diesen Elektronen herausgelöst. Die Geschwindigkeit (und damit die kinetische Energie) dieser Elektronen ist nicht abhängig von der Intensität der einfallenden Strahlung, sondern nur von deren Frequenz. Die maximale kinetische Energie der Fotoelektronen ist nach der Einstein-Relation $E_{kin} = \hbar\omega - A$, wobei A als *Austrittsarbeit* eine materialspezifische Konstante ist. Die durch Licht aus Metallen ausgelösten Fotoelektronen sind Leitungsbandelektronen. Betrachtet man deren Zustände und vernachlässigt ortsabhängige Schwankungen der potentiellen Energie der Elektronen im Festkörper, so lassen sich die energetischen Verhältnisse zum äußeren lichtelektrischen Effekt im *Potentialtopfmodell* darstellen (Abb. 6.8.1–8a). Die Austrittsarbeit ergibt sich aus $A = A_0 - E_F$ (E_F = Fermi-Energie). Aus dem Energiebandspektrum dotierter Halbleiter folgen mehrere Anregungsgleichungen (Abb. 6.8.1–8b). Das Potentialtopfmodell kann nicht angewendet werden.

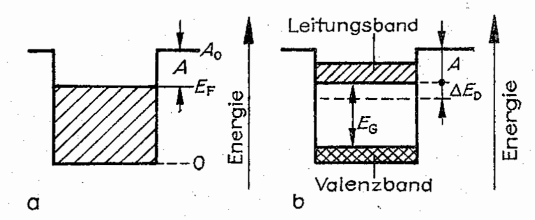


Abb. 6.8.1–8 (a) Potentialtopfmodell für Metallelektronen, (b) Energiebandschema für dotierte Halbleiter

Die langwellige Empfindlichkeitsgrenze ist durch die Bandbreite des Leitungsbandes und die Austrittsarbeit gegeben: $\hbar\omega_g = A$. Mit besetzten Donatorniveaus ist $\hbar\omega_D = A + \Delta E_D$. Für die Emission von Fotoelektronen aus dem Valenzband gilt $\hbar\omega_V = A + E_G$.

Als Material für *Fotokatoden* (in Fotozellen und Sekundärelektronenvervielfachern) im sichtbaren Spektralbereich

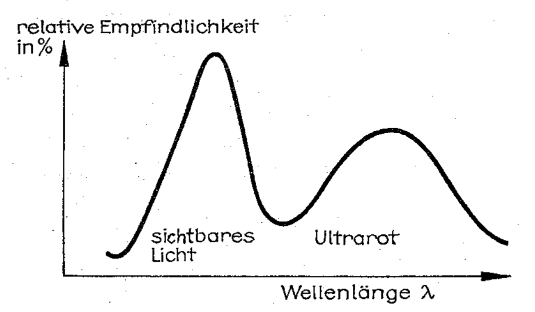


Abb. 6.8.1–9 Fotoelektronenspektrum eines dotierten Halbleiters

werden Störstellenhalbleiter verwendet. Bei Caesiumantimonid (Cs_3Sb) liegt die größte Ausbeute der Fotoelektronen (Empfindlichkeitsmaximum) bei einer Strahlungswellenlänge von 0,4 µm. Durch geeignete Dotierung kann neben der Empfindlichkeit im sichtbaren Spektralbereich eine UR-Empfindlichkeit erzeugt werden. Das ist z. B. bekannt von Caesiumoxid (CsO), in das Silberatome eingelagert sind. Im Energieverteilungsspektrum der Fotoelektronen erscheinen die Elektronen aus dem Valenzband und aus den Donatorniveaus (Abb. 6.8.1–9).

Innerer lichtelektrischer Effekt. Bei Belichtung eines Halbleiters kann es zur Änderung seiner elektrischen Leitfähigkeit kommen. Dieser Effekt, bei dem die Elektronen im Festkörper bleiben, aber vom Valenzband (oder aus Störstellenniveaus) in das Leitungsband gehoben werden (vgl. Abb. 6.8.1–8b), charakterisiert eine material- und wellenlängenabhängige *Fotoleitfähigkeit*. Die Elektronenanregung geschieht durch UV-Strahlung (Übergänge aus dem Valenzband in das Leitungsband) sowie durch Licht und UR-Strahlung (Donatorniveaus-Leitungsband und Valenzband-Leitungsband in schmalbandigen Halbleitern). Bleisulfid (PbS) und Bleitellurid (PbTe) zeigen bereits bei Wellenlängen von 2,3 µm bzw. 6 µm eine ausgeprägte Fotoleitfähigkeit. Alle Fotoleiter sind auch gute elektronische Halbleiter (mit hoher Elektronenbeweglichkeit) und zeigen Lumineszenz. Voraussetzung für einen merklichen Fotostrom ist ein gewisser Verunreinigungsgrad des Basismateriales, der zu mikroskopischen Heteroübergängen mit unterschiedlicher Ladungspolarität (*p*-*n*-Übergänge) führt.

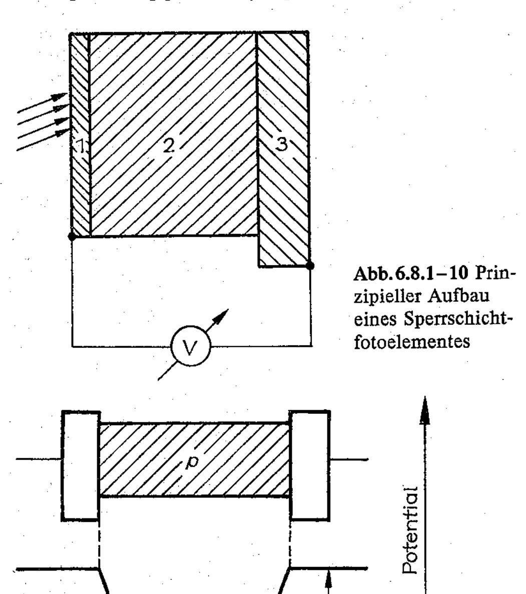


Abb. 6.8.1–10 Prinzipieller Aufbau eines Sperrschichtfotoelementes

Abb. 6.8.1–11 Potentialverlauf am unbelichteten Sperrschichtelement

Sperrschichtfotoeffekt. Er ist die Grundlage für ein *Fotoelement*, eine Batterie, die z. B. als Belichtungsmesser oder als Solarzelle eingesetzt wird. Das Element besteht aus einer Metallunterlage, einer *p*- oder *n*-leitenden Halbleiterschicht und einer dünnen (lichtdurchlässigen) Metallauflage (Abb. 6.8.1–10).

Die Foto-EMK ist vom Halbleitermaterial abhängig. Für Silicium beträgt sie je Element 0,6 V, für Galliumarsenid (GaAs) 0,9 V. Bei materialgleichen Metallkontakten und *p*-leitendem Halbleiter (bei *n*-Leitung kehren sich die Ladungsvorzeichen um) bildet sich an den Grenzflächen Metall-Halbleiter eine Potentialdifferenz ΔU aus (Abb. 6.8.1–11).

Die Akzeptorzustände in der Nähe des Überganges zum Metall werden von Metallelektronen besetzt (Abb. 6.8.1–12). Dieser Ladungsübergang führt zur Ausbildung einer Doppelschicht, die für einen weiteren Elektronenfluß als Sperrschicht wirkt. Das Metall lädt sich positiv auf, der Halbleiter negativ. Durch die Gegeneinanderschaltung der beiden Grenzflächenpotentiale an den Heteroübergängen kompensieren sich die beiden Kontaktpotentiale ($V = 0$ in Abb. 6.8.1–11). Die Kompensation wird zum überwiegenden Teil aufgehoben, wenn man eine Grenzfläche belichtet. Durch die Wechselwirkung mit Licht werden im Halbleiter Elektronen in das Leitungsband gehoben. Sie sind nicht mehr an bestimmte Atome des Halbleiterkristalles gebunden und können den Elektronenmangel im Metall nach Diffusion ausgleichen. Die Elektronenlücken (Defektelektronen, Löcher) schirmen die negative Raumladung im Halbleiter ab. Beide Erscheinungen erhöhen das Potential in der belichteten Grenzflächenregion und führen zur Ausbildung der Potentialdifferenz ΔU_f der Foto-EMK (Abb. 6.8.1–13). Der *Wirkungsgrad* der Fotoelemente für die direkte Umwandlung von Licht in elektrische Energie beträgt etwa 15 % und kann theoretisch einen Maximalwert von 20 % erreichen. Das ist darin begründet, daß es weitere Energieumwandlungen gibt, die nicht zur Potentialerniedrigung auf der belichteten Seite des Elements beitragen: Anregung von Exzitonen, von Fremdatomen und die Energieaufnahme durch freie Elektronen.

Elektrooptischer Effekt. Ein externes elektrisches Feld, das an einen Kristall angelegt wird, kann zu einer Änderung der Brechzahl n führen. Bei $\Delta n \sim E^2$ ist dieser Effekt als *Kerr-Effekt* bekannt geworden, bei $\Delta n \sim E$ als *Pockels-Effekt*.

Piezooptischer Effekt. Er beschreibt die Änderung von n durch mechanische Spannungen, der *Franz-Keldysch-Effekt* die Verschiebung der Absorptionskante durch ein elektrisches Feld. Diese Effekte spielen eine Rolle beim Bau schneller Lichtmodulatoren durch Intensitätsmodulation und in Zukunft in Bildwiedergabegeräten unter der Voraussetzung, daß zweidimensionale digitale Ablenksysteme zur Verfügung stehen.

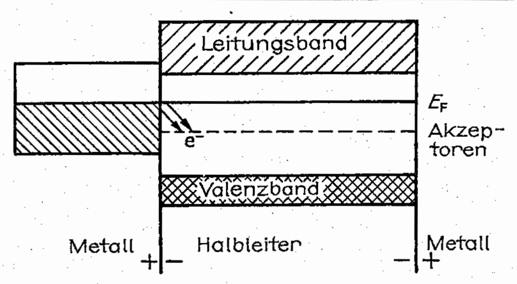


Abb. 6.8.1–12 Energieschema für ein unbelichtetes Sperrschichtelement

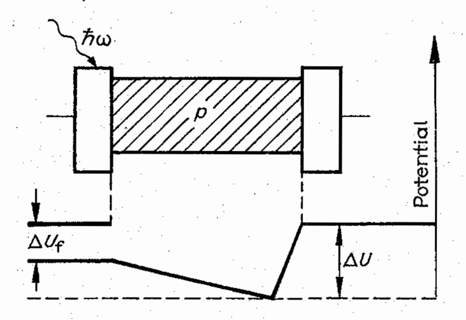


Abb. 6.8.1–13 Potentialverlauf am belichteten Sperrschichtelement

6.8.1.6. Strahlungswechselwirkung mit freien Elektronen. Die elektrische Leitfähigkeit γ beeinflußt die Ausbreitung elektromagnetischer Wellen im Festkörper. Für das sichtbare Licht werden die Leiter undurchsichtig. Die *Eindringtiefe* der Strahlung, nach der die Intensität der Strahlung an der Oberfläche auf das 1/e-fache (e = Basis des natürlichen Logarithmus) reduziert wird, beträgt für gut leitende Metalle $\approx 10^{-2}$ mm. Die Energieverluste des Strahlungsfeldes gehen über in Polarisations- und Elektronenenergie. Die Verlustenergien ergeben sich aus dem frequenzabhängigen Verhalten der komplexen Dielektrizitätsfunktion $\varepsilon(\omega) = \varepsilon_1(\omega) + i\varepsilon_2(\omega)$. Deren Verhalten ist die Antwortfunktion des nichtmagnetischen Festkörpers auf eine elektromagnetische Welle der Frequenz ω. Zwischen ε_1 und ε_2 bestehen feste Beziehungen, die *Kramers-Kronig-Relationen*. In Abhängigkeit von den Eigenfrequenzen ω_i des Festkörpers gilt:

$$\varepsilon_1(\omega_i) = \varepsilon_1(\infty) + (2/\pi) \int_0^\infty \omega\varepsilon_2(\omega)\mathrm{d}\omega/(\omega^2 - \omega_i^2) \quad \text{und}$$

$$\varepsilon_2(\omega_i) = (2\omega_i/\pi) \int_0^\infty \varepsilon_1(\omega)\mathrm{d}\omega/(\omega_i^2 - \omega^2).$$

ε_1 ist bei jeder beliebigen Eigenfrequenz ω_i bekannt, wenn ε_2 im gesamten Spektralbereich vorliegt. $\varepsilon_1(\infty)$ ist die *optische Dielektrizitätskonstante*. Sie gilt für $\omega \to \infty$ und beschreibt nur die elektronische Polarisation. Die *statische Dielektrizitätskonstante* $\varepsilon_1(0) = \varepsilon_1(\infty) + \sum_i S_i$ hingegen berücksichtigt außerdem die Beiträge der i-Gitterschwingungen zur Polarisation. Mit dem Brechungsindex n und dem Absorptionsindex $\varkappa$ ist $\varepsilon(\omega) = [n(\omega) + i\varkappa(\omega)]^2$. Polarisation und Absorption ergeben sich aus $\varepsilon_1(\omega) = n^2 - \varkappa^2$ bzw. $\varepsilon_2(\omega) = 2n\varkappa$. Ohne Berücksichtigung der Interbandübergänge ist die Frequenzabhängigkeit von ε_1: $\varepsilon_1 = A - B/\omega^2$. Dabei berechnen sich A und B aus $A = 1 + N\alpha_{\text{ges}}/\varepsilon_0$ bzw. $B = ne^2/(\varepsilon_0 m_{\text{eff}}) = \omega_p^2$. N ist die Zahl der Atome/cm³ und $\alpha_{\text{ges}} = \alpha_{\text{Atomrümpfe}} + \alpha_{\text{Val. el.}} + \alpha_{\text{o. Polaris.}}$ die Gesamtheit aller Polarisierungsmöglichkeiten. B ist das Quadrat der Plasmafrequenz ω_p. Polarisationsverluste durch Strahlungsanregung von Gitterschwingungen sind an die Bedingung $\varepsilon_1 \gtrsim 1$ geknüpft. Sie ist für $\omega \gtrsim \omega_p$ erfüllt. Bei $\omega < \omega_p$ wird die elektromagnetische Strahlung total reflektiert (vgl. 6.8.1.4. und Abb. 6.8.1–6). Die *Absorptionsverluste* bestehen aus drei Anteilen: $\varepsilon_2 = \varepsilon_2$ (freie Elektr.) + ε_2 (Plasmonen) + ε_2 (IB). Der erste Summand steht für die Absorption durch freie Elektronen (Drude-Absorption), der zweite für die Absorption durch Plasmonenanregung, der dritte Summand beschreibt direkte und indirekte Interbandübergänge (vgl. 6.8.1.1.). Für die Frequenz der Plasmaschwingungen ω_p im Kristallvolumen (die Frequenz der *Oberflächenplasmonen* für Metalle ist $\omega_p^{\text{OF}} = \omega_p/\sqrt{2}$) hat die *Energieverlustfunktion* $-\mathrm{Im}\,\varepsilon^{-1}(\omega) = \varepsilon_2/(\varepsilon_1^2 + \varepsilon_2^2)$ ihren größten Wert. Er folgt aus $\varepsilon_1 = 0$, d.h. aus $n = \varkappa$.

Ellipsometrie dient der Bestimmung von n und $\varkappa$. Bei diesem Experiment fällt eine linear polarisierte, ebene monochromatische Lichtwelle auf eine Kristalloberfläche. Nach der Reflexion ist die Welle elliptisch polarisiert. Die Analyse des elliptischen Schwingungszustandes liefert die optischen Konstanten (Abb. 6.8.1–14).

Aus den Fresnelschen Formeln (vgl. 6.7.) ergeben sie

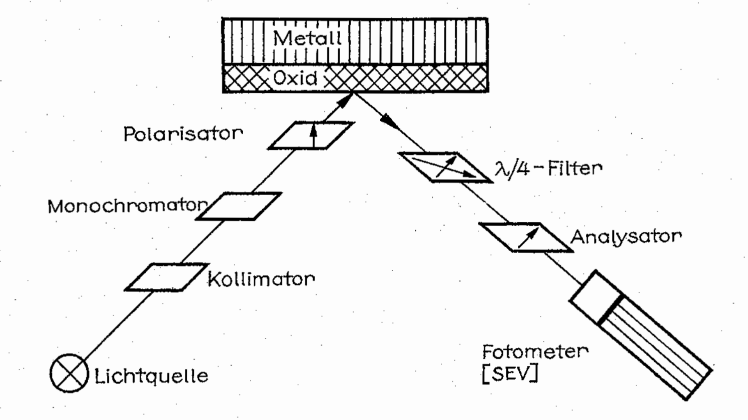


Abb. 6.8.1–14 Schema einer Versuchsanordnung zur Ellipsometrie

sich zu $n = [\sin\varphi \tan\varphi \cos 2\varrho]/[1 + \sin 2\varrho \cos\Delta]$ und $\varkappa = \sin\Delta \tan 2\varrho$.
Das elliptisch polarisierte Licht besteht aus zwei Komponenten $\boldsymbol{E}_p$ und $\boldsymbol{E}_s$. Diese Feldvektoren sind parallel und senkrecht zur Einfallsebene polarisiert. Das Verhältnis ihrer Beträge ist $\tan\varrho = |\boldsymbol{E}_p|/|\boldsymbol{E}_s| = (I_p/I_s)^{1/2}$ (I = Strahlungsintensitäten). φ ist der Einfallswinkel der linear polarisierten Welle und Δ die Phasendifferenz zwischen $\boldsymbol{E}_p$ und $\boldsymbol{E}_s$. Für den *Haupteinfallswinkel* φ_H ist $\Delta = \pi/2$ und $|\boldsymbol{E}_s|$ hat ein Minimum. Die optischen Konstanten reagieren sehr empfindlich auf Veränderungen der Oberfläche. Geringste Kontamination der Oberfläche verändert deren Werte. Die Ellipsometrie ist so empfindlich, daß man auf 10^3 Atome an der Oberfläche eines Wirtsgitters 1 Verunreinigungsatom nachweisen kann.

6.8.1.7. Wechselwirkung mit ultravioletter Strahlung. Elektronenschwingungsspektren. Viele Kristalle haben im UV-Strahlungsbereich scharfe Emissions- und Absorptionsspektren. Sie werden durch elektronische Übergänge bestimmt. Die Anregung erfolgt aus Leitungs- bzw. Valenzbändern oder aus Störstellenniveaus in unbesetzte Zustände im oder außerhalb des Festkörpers. Die Absorptions- bzw. Emissionsspektren sind Abbilder der obersten besetzten bzw. der darüber liegenden unbesetzten Energieniveaus, moduliert durch spektrometrische *Auswahlregeln*, die erlaubte von verbotenen Übergängen trennen. Für die erlaubten Übergänge ist $\Delta j = 0$; ± 1 (LS-Kopplung) und $\Delta J = 0$; ± 1 (jj-Kopplung), wobei $0 \to 0$-Übergänge verboten sind. Die Auswahlregeln gelten nicht streng, sie können durch Störungen (z. B. hohe elektrische oder magnetische Felder) aufgehoben werden. Die elektronischen Zustände im Festkörper können mit Gitterschwingungen gekoppelt sein. Diese Modulation verbreitert die scharfen UV-Spektren durch *Seitenbänder*, die charakteristisch sind für die Erzeugung bzw. Vernichtung von Phononen. Aus der Feinstruktur der Seitenbänder lassen sich die Frequenzen der Gitterschwingungen des Kristallgitters ermitteln. Die elektronischen und Kristallgittereigenschaften der Festkörper sind bestimmend für die Transmission, Absorption und Reflexion elektromagnetischer Strahlung. Neben den Primäreffekten der Wechselwirkung werden auch die Sekundäreffekte (Lumineszenz, Fotoeffekte u. a.) beeinflußt. Besondere Bedeutung in der Festkörperspektrometrie hat die Untersuchung von Fotoelektronen erlangt, die durch Strahlungswechselwirkung aus dem Festkörper herausgelöst werden. Als anregende Strahlung verwendet die *Fotoelektronenspektrometrie* ultraviolette Strahlung (für UPS) und Röntgen-Strahlung (für XPS, vgl. 6.8.1.8.).

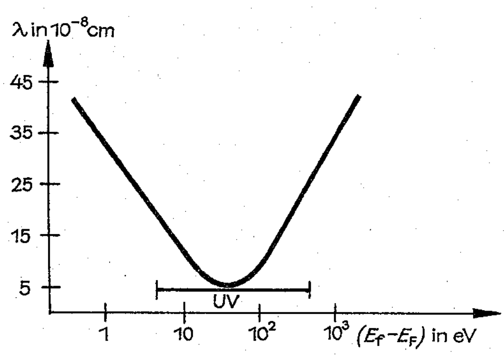


Abb. 6.8.1–15 Mittlere freie Weglänge der Fotoelektronen im Festkörper als Funktion der Elektronenenergie (E_F = Fermi-Energie)

Die *Fotoelektronenemission* zerfällt nach einer Modellvorstellung in drei aufeinander folgende, voneinander unabhängige Ereignisse: die optische Anregung des Elektrons am Atom, die zur Loslösung vom Atom führt, Transport des angeregten Elektrons zur Oberfläche und Überwindung der Barriere des Oberflächenpotentials der Festkörperoberfläche. Die Spektrometrie der nun freien Elektronen umfaßt die Bestimmung der *Energie-* und *Winkelverteilung* und der *Quantenausbeute*. Die Verteilung informiert über den Anteil $N(E)$ der Elektronen, die eine Energie im Intervall $E \dots E + \mathrm{d}E$ aufweisen. Die Ausbeute ist die Zahl der emittierten Elektronen je absorbiertem Photon. Sie ist das Integral der Energieverteilung über alle Energien, die für Fotoelektronen nach den Anregungsbedingungen möglich sind: $Y(E, \hbar\omega) = \int_0^{E_{max}} N(E)\mathrm{d}E$. Für die *Ionisierungswahrscheinlichkeit* des Festkörperatoms gilt: $\delta(E, \hbar\omega)\mathrm{d}E\,\mathrm{d}x = \sigma(E, \hbar\omega)\,[1 - R(\hbar\omega)]\exp[-K(\hbar\omega)x]\mathrm{d}E\,\mathrm{d}x$. Dieses Atom befindet sich im Abstand x von der Oberfläche. K ist der Absorptionskoeffizient, R das optische Reflexionsvermögen an der Festkörperoberfläche. Auf dem Weg zur Oberfläche kann das Fotoelektron Energie verlieren. Die wichtigsten Verlustmechanismen sind Kollision mit anderen Elektronen, Plasmonenanregung und Elektron-Phonon-Wechselwirkung. Die Austrittstiefe für Fotoelektronen im Festkörper ist begrenzt. Die Wahrscheinlichkeit, daß ein Elektron der Energie E_f die Oberfläche ohne große Energieverluste erreicht, ist $Q(E_f) = \exp[-x/\lambda(E_f)]$. $\lambda(E_f)$ ist die *mittlere freie Weglänge* der Fotoelektronen im Festkörper. Ihre Energieabhängigkeit zeigt über eine Vielzahl von Elementen und Verbindungen gemittelt einen ähnlichen Verlauf (Abb. 6.8.1–15).

Bei der Anregung mit UV-Strahlung kommen die Elektronen aus einer nur sehr dünnen Oberflächenschicht ($\approx 5 \cdot 10^{-8}$ cm). Die Oberflächenbarriere überwindet das Fotoelektron um so leichter, je größer seine kinetische Energie E_f, je geringer seine Bindungsenergie E_B ist. Die Gesamtwahrscheinlichkeit des Elektronenaustritts aus einer beliebigen Tiefe x je absorbiertem Photon ($R = 0$) ist: $P(E, \hbar\omega) = \sigma B/(K + 1/\lambda)$. $B(E)$ ist die Wahrscheinlichkeit zur Überwindung der Oberflächenbarriere und σ die differentielle optische Absorptionskonstante, die durch die Übergangswahrscheinlichkeit der Elektronen von den besetzten Anfangszuständen in die unbesetzten Endzustände bestimmt ist. Übergangswahrscheinlichkeit und mittlere freie Weglänge sind von der Frequenz der anregenden Strahlung abhängig.

Die Frequenzänderung zur Optimierung des Spektrometers läßt sich am besten bewerkstelligen mit einer durchstimmbaren UV-Strahlungsquelle. Diese ist in Form des *Elektronensynchrotrons* gegeben. Das Synchrotron (vgl. 5.4.) liefert aufgrund relativistischer Effekte ein Strahlungskontinuum, das vom Licht bis zum harten Röntgen-Strahlgebiet reicht. Die Strahlung ist von hoher Intensität (10^3 mal größer als die anderer Strahlungsquellen), stark gebündelt und polarisiert. Bei extrem kurzen Meßzeiten gewinnt man zuverlässige Aussagen zur Ladungsverteilung im Festkörper.

Die Untersuchung von Festkörperoberflächen mit Hilfe elektromagnetischer Strahlung hat gegenüber der mit Teilchenbeschuß (vgl. 6.8.2.) einen erheblichen Vorteil. Bei der Bestrahlung gibt es keine Veränderungen auf den zu testenden Oberflächen. Das kann hingegen bei Teilchenbeschuß

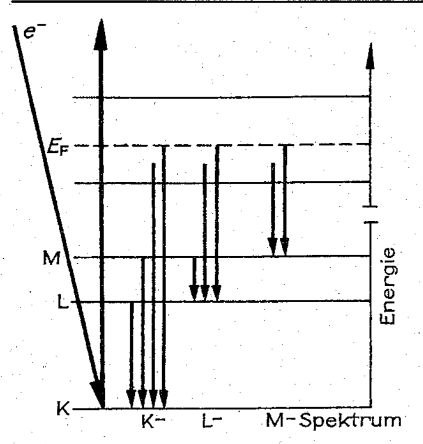


Abb. 6.8.1–16 Entstehung des Röntgen-Emissionsspektrums (Linienspektrum)

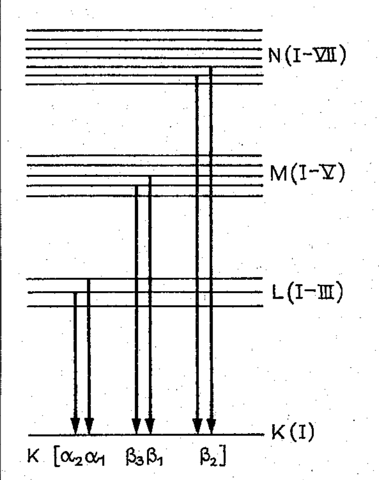


Abb. 6.8.1–17 Feinstruktur des Röntgen-K-Emissionsspektrums

passieren, wie die *elektronenstrahlstimulierte Desorption*, d. h. die Abgabe von Oberflächenatomen in das Vakuum beim Beschuß der Festkörperoberfläche mit Elektronen, zeigt.

6.8.1.8. Wechselwirkung mit Röntgen- und γ-Strahlung. Die Röntgen-Strahlung hat Wellenlängen, deren Größe den Atomabständen im kristallinen und amorphen Festkörper entsprechen, $\lambda \approx 10^{-8}$ cm. Deshalb kommt es bei der Wechselwirkung zwischen dieser Strahlung und der kondensierten Materie zur Röntgenstrahlbeugung. Sie unterscheidet sich für kristalline (dazu gehören etwa 95 % der anorganischen Festkörper) und amorphe Substanzen deutlich voneinander. Die Beugungseffekte können zur Aufklärung der Struktur der Festkörper verwendet werden.

Röntgen-Strahlemission. Ein Festkörper emittiert Röntgen-Strahlen, wenn er mit Elektronen hoher kinetischer Energie beschossen wird. Diese schlagen aus kernnahen Elektronenschalen (K, L, M) Elektronen heraus, die einen unbesetzten Zustand im Target besetzen oder als Fotoelektron emittiert werden. Die entstandenen Elektronenlücken werden durch Elektronenübergänge aus energiereicheren Schalen bzw. Bändern aufgefüllt. Dabei entsteht eine *charakteristische Röntgen-Strahlung* in Form von Linienspektren (Abb. 6.8.1–16). Alle Linien, deren Endzustand die K-Schale ist, gehören zum *K-Emissionsspektrum*. Die Emissionslinien bilden ein für den jeweiligen Festkörper charakteristisches Spektrum. Die Frequenzen ω sind durch das *Moseleysche Gesetz* $\omega = A(Z - B)^2$ (mit Z als Ordnungszahl des Elementes, A und B sind Konstante) bestimmt. Für das K-Spektrum gilt $\omega = (3/2)\pi R(Z-1)^2$ (R = Rydberg-Konstante). Das L- bzw. M-Spektrum entsteht durch Auffüllen der in den entsprechenden Schalen gebildeten Elektronenleerstellen.

Spektrenfeinstruktur. Die Elektronenschale mit der Hauptquantenzahl n besteht aus $2n - 1$ energetisch eng benachbarten Zuständen. Das spiegelt sich wider in einer Feinstruktur der Röntgen-Emissionsspektren. Die möglichen Elektronenübergänge werden wie bei den optischen Spektren durch Auswahlregeln gesteuert. So fehlen im Röntgen-Spektrum der K-Emission die Linien, die den Übergängen $2S_{1/2} - 1S_{1/2}$ oder $D \rightarrow S(\Delta j > 1)$ (Abb. 6.8.1–17) entsprechen.

Geht ein Valenzelektron auf ein freies Rumpfniveau über, so kann das von verschiedenen energetischen Lagen aus vor sich gehen. Das Valenzband wird in einem *Röntgen-Emissionsband* abgebildet. Die Intensitätsverteilung der Emissionsbande wird durch die Zustandsdichte $N(E)$ bestimmt, wobei $N(E)\mathrm{d}E$ die Zahl der Elektronen im Intervall $E \ldots E + \mathrm{d}E$ darstellt.

Röntgen-Bremsstrahlung. Wenn die beschleunigten Elektronen, die zur Emission der charakteristischen Strahlung führen, auf den Festkörper auftreffen, werden sie zunächst von den äußeren Elektronen der Atome des Targets infolge der Coulomb-Wechselwirkung abgebremst. Dieser Energieaustausch führt zur Emission kontinuierlicher („weißer") Röntgen-Strahlung, dem Bremskontinuum. Es ist insofern materialspezifisch, als die abgestrahlte integrale Intensität $I \sim Zu^2 i$ ist, wobei u die Beschleunigungsspannung und i der Elektronenstrom der strahlungsauslösenden Elektronen sind. Dem Kontinuum ist das charakteristische Spektrum überlagert (Abb. 6.8.1–18). Die *kurzwellige Kante* folgt aus dem Energiesatz zu $\lambda_{\min}[10^{-8}\,\mathrm{cm}] = 12{,}34/u\,[\mathrm{kV}]$.

Röntgen-Strahlabsorption. Den kurzen Wellenlängen der Röntgen-Strahlung entspricht eine hohe Strahlungsenergie. Für $\lambda \approx 2 \cdot 10^{-8}$ cm ist sie ≈ 10 keV. Aufgrund dieser hohen

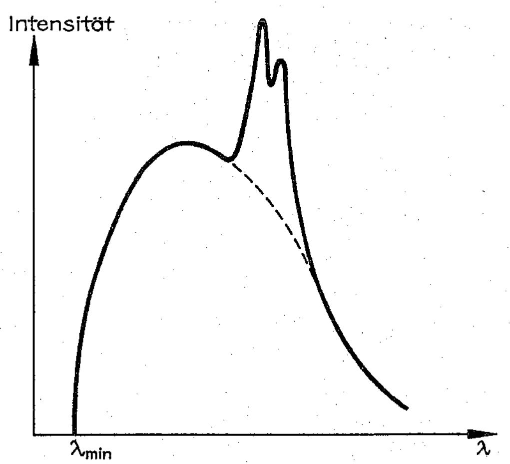


Abb. 6.8.1–18 Überlagerung von Bremsspektrum und charakteristischem Spektrum im Röntgen-Emissionsspektrum

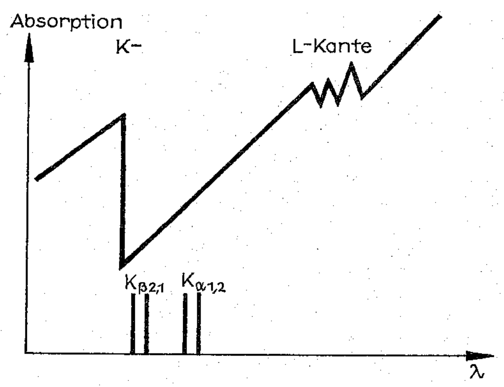


Abb. 6.8.1–19 Röntgen-Absorptionsspektrum

Energie ist das Eindringvermögen der Strahlung in feste Körper wesentlich größer als für das Licht oder für UV-Strahlung. Im Unterschied zu optischen Spektren, bei denen es Linienspektren in Emission und Absorption gibt, hat das Röntgen-Absorptionsspektrum keine Linien. Es ist ein kontinuierliches Spektrum (Abb. 6.8.1–19).

Die Intensitätsverminderung beim Strahlungsdurchgang durch den Festkörper geschieht durch Absorption und durch Streuung. Demnach gilt Schwächung (μ) = Absorption (τ) + Streuung (σ). Dabei ist $\tau \sim Z^4\lambda^3$ und $\tau > \sigma$. Das Absorptionsspektrum besteht aus *Absorptionskanten*, die sich aus den Elektronenniveaus (K, L_{I-III} usw.) ergeben. Die Wellenlängen der den Kanten entsprechenden Emissionslinien liegen aus energetischen Gründen oberhalb derselben (Abb. 6.8.1–19). Die Multiplizität der Energieterme mit $n > 1$ verursacht eine Kantenaufspaltung (*Kantenfeinstruktur*). Für den Festkörper gibt es eine *ausgebreitete Feinstruktur* der Röntgen-Absorptionskanten, die sich über einen Energiebereich von einigen 100 eV erstreckt (vgl. EXAFS). Die Absorptionskanten weicher Röntgen-Strahlung (Energie $\lesssim 1$ keV) verändern ihre Lage in Abhängigkeit vom Bindungszustand der Strahlung emittierenden Atome.

Röntgen-Strukturanalyse. Ein kristalliner Festkörper besteht aus dreidimensional angeordneten Atomen, Ionen bzw. Atomgruppen (vgl. 6.1.). Durch diese Bausteine eines Kristallgitters lassen sich in verschiedene Richtungen äquidistante Ebenenscharen, sogenannte *Netzebenen*, legen. Der *Netzebenenabstand d* in der Schar bestimmt sich aus der Orientierung der Netzebenen relativ zu einem kristallfesten Koordinatensystem, das durch die Wahl der Elementarzelle (vgl. 6.1.) bestimmt ist. Die Beugung der Röntgen-Strahlen der Wellenlänge λ ist beschreibbar als Reflexion an atomaren Punktladungen (Abb. 6.8.1–20). Die Streurichtungen Θ leiten sich aus der *Braggschen Gleichung* ab: $n\lambda = 2d \sin \Theta$.

Die Bragg-Reflexion unterscheidet sich von der spiegelnden (kontinuierlichen) Reflexion durch die Winkel- und Wellenlängenabhängigkeit als Ergebnis der gitterperiodischen Anordnung der Streuzentren. Mit der Reflexionsordnung n (= 1; 2; 3; ...) ergeben sich nur unter bestimmten Glanzwinkeln Θ bei bestimmten Wellenlängen λ abgebeugte Strahlen. Für diese ist der Gangunterschied zwischen den an den Netzebenen (1) und (2) reflektierten Strahlen $2d \sin \Theta$ ein ganzzahliges Vielfaches von λ. Die sich überlagernden atomaren Einzelwellen sind dagegen in gleicher Phase. Der Netzebenenabstand d kann gemäß der Bragg-Gleichung aus einem Streuexperiment bestimmt werden. Für eine rechtwinklige Elementarzelle ($\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$) ist für die Ebenenschar der Millerschen Indizes (hkl) (vgl. 6.1.2.2.) der Netzebenenabstand $d = [(h/a_1)^2 + (k/a_2)^2 + (l/a_3)^2]^{-1/2}$. Die Streuintensität in die Richtung Θ (vgl. Abb. 6.8.1–20) entspricht dem Amplitudenquadrat der Streuwelle. Die einfallende Röntgen-Welle (Kreisfrequenz ω, Wellenzahlvektor $\boldsymbol{k} = (2\pi/\lambda) \cdot \boldsymbol{e}$; $\boldsymbol{e}$ = Einheitsvektor der Wellennormalen) wird an den atomaren Streuzentren (Ortsvektoren $\boldsymbol{R} = m\boldsymbol{a}_1 + n\boldsymbol{a}_2 + p\boldsymbol{a}_3$; m, n, p sind ganze Zahlen) gestreut. Die Einzelstreuwellen mit den Phasenfaktoren $\exp[-i\boldsymbol{R}(\boldsymbol{k}' - \boldsymbol{k})]$ ($\boldsymbol{k}'$ = Wellenzahlvektor der reflektierten Welle) summiert ergeben die *Streuamplitude A* $= \sum_{mnp}^{LLL} \exp[-i\boldsymbol{R}(\boldsymbol{k}' - \boldsymbol{k})]$. Sie hat einen Maximalwert von L^3, wenn $\boldsymbol{R}(\boldsymbol{k}' - \boldsymbol{k})$ ein ganzzahliges Vielfaches von 2π ist. Daraus folgen die *Laue-Gleichungen*: $\boldsymbol{a}_1(\boldsymbol{k}' - \boldsymbol{k}) = 2\pi h$; $\boldsymbol{a}_2(\boldsymbol{k}' - \boldsymbol{k}) = 2\pi k$; $\boldsymbol{a}_3(\boldsymbol{k}' - \boldsymbol{k}) = 2\pi l$. Mit den Laue-Gleichungen ist die Streuamplitude $A = \sum_{mnp}^{LLL} \exp[-2\pi i(mh + nk + pl)]$.

Es ist vorteilhaft, den Differenzvektor der Wellenzahlen $\Delta\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k}$ in einem *reziproken Raum* (Fourier-Raum) mit den Basisvektoren $\boldsymbol{b}_i$ (vgl. 6.3.1.) so darzustellen, daß die Millerschen Indizes als Faktoren erscheinen, $\Delta\boldsymbol{k} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3$. Jeder Punkt des reziproken Gitters, mit dem reziproken Gittervektor $\boldsymbol{G} = h\boldsymbol{b}_1 + k\boldsymbol{b}_2 + l\boldsymbol{b}_3$, beschreibt eine Netzebenenschar (hkl). Die Streuamplitude hat ein Maximum, wenn $\Delta\boldsymbol{k}$ ein Vektor des reziproken Gitters ist. Die *Ewald-Kugel* im reziproken Gitter informiert über die möglichen Reflexionsebenen (hkl) (Abb. 6.8.1–21).

Für parallele Röntgen-Strahlen sind die Abmessungen der

Abb. 6.8.1–20 Bragg-Reflexion an den Netzebenen eines Kristallgitters

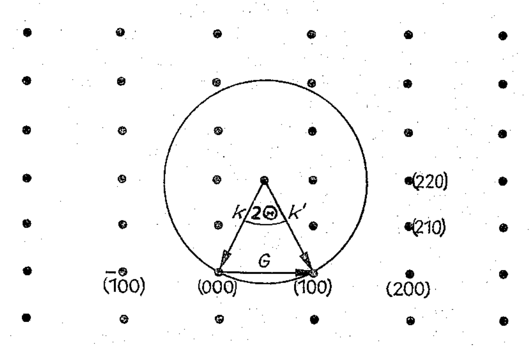


Abb. 6.8.1–21 Ewald-Konstruktion für ein zweidimensionales reziprokes Gitter

Ewald-Kugel so, daß lediglich zwei Punkte des reziproken Gitters die Interferenzbedingung erfüllen (Zweistrahlinterferenz). Befinden sich in der Elementarzelle (a_1, a_2, a_3) mehrere (s) Atome, so können deren relative Lagen in der Einheitszelle (innere Koordinaten $r_j = (x_j, y_j, z_j)$ die gestreute Bragg-Intensität modulieren: $A = A_{\max} S$.

$S = \sum_{j=1}^{s} f_j \exp[-2\pi i (x_j h + y_j k + z_j l)]$ ist die *Strukturamplitude*, S^2 der *Strukturfaktor*. Die *Atomformfaktoren* f_j berücksichtigen das unterschiedliche Streuvermögen der Atome infolge der unterschiedlichen Elektronendichteverteilung um den Atomkern. Die Atomformfaktoren wären gleich der Zahl der Elektronen Z, wenn diese in einer Punktladung (ohne räumliche Ausdehnung) zusammengefaßt wären.

Laue-Verfahren. Bei diesem Verfahren der Strukturanalyse wird ein Einkristall (vgl. 6.1.) von polychromatischen Röntgen-Strahlen durchstrahlt. Die für ganz bestimmte Wellenlängen in Bragg-Stellung befindlichen Netzebenenscharen beugen den Primärstrahl ab. Das Beugungsmuster wird auf einem hinter dem Kristall angebrachten Röntgen-Film (oder auf Fotoplatte) fixiert. Fällt die Primärstrahlrichtung mit einer Symmetrieachse des Kristalles zusammen, so zeigt das Laue-Diagramm die der Zähligkeit der Achse entsprechende Symmetrie. Das Laue-Verfahren eignet sich zur Orientierungsbestimmung von einkristallinen Festkörpern.

Beim *Drehkristallverfahren* wird ein Einkristall von einem monochromatischen Röntgen-Strahlbündel durchsetzt. Der Kristall wird um eine Achse gedreht, die senkrecht zur Primärstrahlrichtung orientiert ist. Um den Kristall ist zylinderförmig ein Film angebracht, auf dem man die gebeugten Strahlen registrieren kann. Durch die Drehbewegung des Kristalles kommen verschiedene Netzebenen (*hkl*) in Reflexionsstellung. Die Diagramme sind Schichtliniendiagramme. Aus dem Abstand der Schichtlinien zum Äquator ergibt sich die Gitterkonstante parallel zur Drehrichtung. Mithin können mit dem Verfahren auch Orientierungsbestimmungen vorgenommen werden. Verbesserungen dieser Methode arbeiten mit synchroner Filmverschiebung bzw. Filmdrehung (Verfahren nach WEISSENBERG-BÖHM bzw. SCHIEBOLDT-SAUTER), wodurch mit erhöhtem Auflösungsvermögen eine Indizierung der Reflexe ermöglicht wird, aus denen sich die Schichtlinien zusammensetzen.

Für Strukturuntersuchungen am polykristallinen Material wird das *Debye-Scherrer-Verfahren* verwendet. Die Kristalle werden monochromatisch durchstrahlt. Die Probe wird in einer Zylinderkammer um ihre Längsachse gedreht. Die Röntgen-Reflexe werden auf Film registriert. Aufgrund des polykristallinen Charakters der Probe liegen die gebeugten Strahlen auf einem Kreiskegel mit der Primärstrahlrichtung als Achse. Auf dem Film erscheinen als Interferenzlinien die Schnittlinien der verschiedenen Kreiskegel mit der Filmebene als konzentrische Ringe. Durch Vermessung der Ringdurchmesser auf dem Äquator lassen sich die Abmessungen der Elementarzelle berechnen. Die Anordnung der Debye-Scherrer-Reflexe ist charakteristisch für die verschiedenen Kristallstrukturen (Bravais-Gitter, vgl. 6.1.). Die Intensität der Reflexe wird bestimmt durch die Art der Atome (Ordnungszahl), die die Netzebenen bilden, den Besetzungsgrad der Ebenen und durch Interferenzeffekte (Strukturfaktor). Die Breite der Beugungslinien ist abhängig von der Kristallitgröße. Wird eine bestimmte Größe (Durchmesser 10^{-4} mm) unterschritten, so werden die Linien unschärfer. Die winkelabhängige Intensitätsverteilung eines Bragg-Reflexes, die *Rocking-Kurve*, ist ein Hinweis darauf, daß auch für einen Winkelbereich in der Umgebung des Bragg-Winkels Θ Beugung auftritt. Wenn die Gleichphasigkeit der Einzelwellen, aus denen der Bragg-Reflex gebildet wird, nicht mehr gegeben ist, kommt es zur Verbreiterung der Rocking-Kurve. Sie entsteht, wenn man vom Einkristall zum polykristallinen Material kleiner Kristallite übergeht bzw. wenn die Kristallperfektion durch Defekte, wie Punktdefekte, Stapelfehler, Kleinwinkelkorngrenzen und Zwillingsgrenzen, gestört ist. Der Automatisierung des Meßvorganges dient beim Debye-Scherrer-Verfahren ein bewegliches Zählrohr, mit dessen Hilfe die Intensität der gebeugten Strahlung winkelabhängig gemessen werden kann. Diese Anlagen heißen *Diffraktometer* (Zählrohrgoniometer). Definiert man die Massendichte ϱ mit Hilfe der Elementarzelle des Kristalles, so ist $\varrho = (Ms)/(LV)$ (M = relative Atommasse, L = Loschmidt-Zahl, V = Volumen der Elementarzelle, s = Zahl der Atome bzw. bei Verbindungen Formeleinheiten je Elementarzelle). Bei bekannter Dichte und dem mittels Debye-Scherrer Analyse gewonnenen Elementarzellenvolumen ist die Zahl der Atome je Elementarzelle zugänglich. Die Strukturaufklärung ist vollständig nach der Ermittlung der inneren Koordinaten r_j in der Elementarzelle. Sie sind bestimmend für die Phasenwinkel der Streuwellen. Aus den Intensitäten der Interferenzlinien können keine Schlußfolgerungen auf die Phasenwinkel gezogen werden. Die Intensitäten liefern nur Werte für das Quadrat der Strukturamplitude S, so daß deren Vorzeichen nicht bekannt ist. Abhilfe schafft die Vorgabe theoretischer Strukturmodelle, deren Strukturfaktor S^2 mit den Intensitäten der Beugungsreflexe korreliert wird. Durch sukzessive Verbesserung der Modellparameter gelangt man auch bei komplizierteren Kristallstrukturen ($s > 50$) mit Hilfe der Computersimulation zu befriedigenden Resultaten.

Debye-Waller-Faktor. Die Temperaturerhöhung des Kristalles führt zu einer Abnahme der Intensität der Röntgen-Beugungsreflexe, wobei die Rocking-Kurve nicht breiter wird. Allerdings erhöht sich der diffuse Streuuntergrund des Beugungsspektrums. Die Temperatur verändert die Position der Atome zeitabhängig um die Gleichgewichtslage $\boldsymbol{R}_0$: $\boldsymbol{R}(t) = \boldsymbol{R}_0 + \boldsymbol{u}(t)$. Das Temperaturmittel der gestreuten Amplitude ist $\langle A \rangle = \langle A_0[\exp(-i\boldsymbol{uG})]\rangle$. Aus einer Reihenentwicklung des Exponentialfaktors bis zu Gliedern 2. Ordnung folgt für die Intensität des Beugungsreflexes $I = I_0 \exp\left[-\frac{1}{3}\langle \boldsymbol{u}^2\rangle G^2\right]$.

Der Exponentialfaktor ist der Debye-Waller-Faktor.

Ist die Fernordnung im Festkörper gestört, so führt die Überlagerung der von den Hüllenelektronen emittierten Einzelwellen zu einer *diffusen Streuung* der eingestrahlten Primärwelle. Sie zeigt keine ausgeprägten Interferenzmaxima, die Winkelabhängigkeit der Intensität ist eine mit wachsendem Θ monoton fallende Funktion. Das Übergangsstadium von der Beugung zur diffusen Streuung kann mit einer Verkleinerung der Kristallite experimentell erfaßt werden.

Dynamische Theorie. Bei der Beschreibung der Bragg-Reflexion wurde bisher nur ein Streuvorgang, die Einfachstreuung, berücksichtigt. Bei größeren Kristalldimensionen ($\gtrsim 10$ µm) und nahezu idealer Kristallperfektion müssen Effekte der Mehrfachstreuung beachtet werden, denn eine reflektierte Welle bildet durch Bragg-Reflexion wieder eine „einfallende" Welle, die wiederum die Bragg-Bedingung erfüllt. Die Einfachstreuung, die durch die kinematische (geometrische) Theorie erfaßt wird, beugt aus der einfallenden Primärwelle eine Sekundärwelle. Wechselwirkungen zwischen einfallender und reflektierter Welle werden von

der kinematischen Theorie ausgeschlossen. Im perfekten Kristall ausreichender Abmessungen kommt es aber zu meßbaren Überlagerungseffekten. Die Intensitätsverteilung wird durch Interferenzeffekte (Kossel-Interferenzen) moduliert. Es entstehen Intensitätsschwankungen, die sogenannten Pendellösungsstreifen, deren Abstand L Auskunft gibt darüber, daß bei Kristallabmessungen von etwa 0,1 L die Intensitäten der einfallenden und der reflektierten Wellen ähnlich sind und durch Interferenz zu einer nachweisbaren Intensitätsmodulation im Beugungsdiagramm führen. Für Kristalldimensionen unterhalb von 0,1 L bzw. für gestörte Kristalle (Mosaikkristalle) ist die kinematische Theorie eine ausreichende Näherung. Die dynamischen Beugungserscheinungen sind geeignet, geringste Abweichungen von der idealen Kristallperfektion nachzuweisen. Die zu erwartenden gestörten Interferenzstreifensysteme können mit Hilfe computersimulierter Rechnungen auf der Grundlage der Beugungstheorie an gestörten Kristallen vorhergesagt und mit dem Experiment verglichen werden.

Röntgen-Spektrometrie. Röntgen-Strahlung erzeugt bei Wechselwirkung mit dem Festkörper durch Primäranregung ein Elektronenloch in einem Rumpfelektronenniveau. Der unbesetzte Elektronenplatz wird durch ein Elektron aus dem Valenzband aufgefüllt. Die Energiedifferenz des elektronischen Überganges $E_{nk} - E_c$ wird als Strahlungsenergie in Form eines *Röntgen-Emissionsspektrums* ausgesandt (Abb. 6.8.1–22). Die Intensitätsverteilung der Emissionsstrahlung informiert über die mit Elektronen besetzten Zustände der Energien E_{nk} des Valenzbandes (n = Bandindex, $\boldsymbol{k}$ = Wellenzahlvektor der Elektronen im Band als Maß für deren Impuls $\hbar\boldsymbol{k}$). An der unteren Kante des Bandes ist $\boldsymbol{k} = 0$, für den obersten besetzten Zustand gilt in der Näherung freier Elektronen $k_F = (3\pi^2 N^*/V)^{1/3}$ (N^* = Anzahl der Valenzelektronen im Volumen V des Festkörpers). Aus dem energieabhängigen Intensitätsverlauf der Röntgenemission erhält man Aussagen zur elektronischen Zustandsdichte $N(E)$ des Valenzbandes. Das Emissionsspektrum (Abb. 6.8.1–23) entspricht nicht direkt dem Verlauf der Energieabhängigkeit der elektronischen Zustandsdichte. Es ist moduliert mit der Übergangswahrscheinlichkeit von Ausgangs- und Endzustand, $P(E, \boldsymbol{k}) = |\langle\psi_c|\boldsymbol{er}|\psi_{nk}\rangle|^2$. Der Übergangsoperator ($\boldsymbol{er}$) für den Dipolübergang (vgl. 2.6.3.) setzt sich zusammen aus dem Polarisationsvektor $\boldsymbol{e}$ und dem Ortsoperator $\boldsymbol{r}$ (vgl. 4.). Der Informationsgehalt der Spektren ist höher für polarisierte Röntgen-Strahlung von Einkristallen.

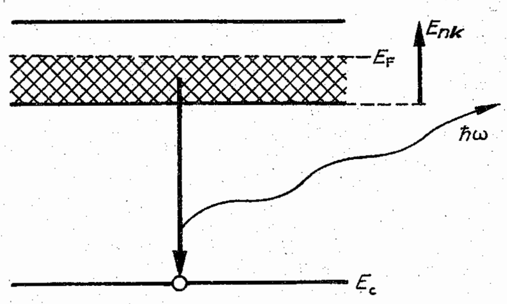


Abb. 6.8.1–22 Valenzbandspektrum in Röntgen-Emission

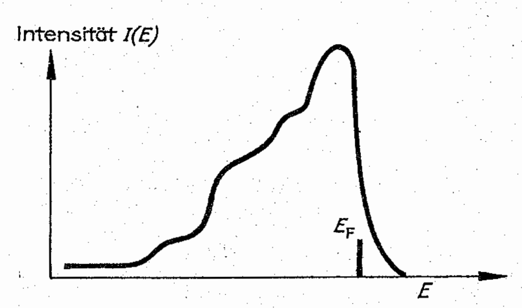


Abb. 6.8.1–23 Röntgen-Emissionsspektrum eines Festkörpers

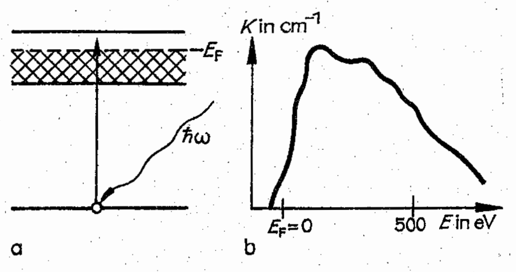


Abb. 6.8.1–24 (a) Valenzbandspektrum in Röntgen-Absorption, (b) Röntgen-Absorptionsspektrum eines Festkörpers

Das *Röntgen-Absorptionsspektrum* beinhaltet die Energieabhängigkeit des Absorptionskoeffizienten $K(E)$. Der Verlauf wird beeinflußt durch die Übergangswahrscheinlichkeit in die oberhalb des Fermi-Niveaus E_F befindlichen unbesetzten Zustände und deren elektronischer Zustandsdichte (Abb. 6.8.1–24). Der Absorptionskoeffizient wächst an der Kante steil an und fällt in komplizierter Weise (Oszillationen) monoton ab für Energien bis zur Größenordnung von 1 keV über der Absorptionskante. Der Verlauf bis zu 30 eV nach der Kante signalisiert Veränderungen der Elektronenwellenfunktionen des angeregten Zustandes durch das Coulomb-Feld des angeregten Atoms. Außerdem spielen die elektronische Bandstruktur und Vielteilchenwechselwirkungen eine Rolle. Der Verlauf des Absorptionskoeffizienten für Energien mit einer Kantendistanz >30 eV heißt *EXAFS* (*E*xtended *X*-ray-*A*bsorption *F*ine *S*tructure). EXAFS wird in Kristallen und amorphen Substanzen beobachtet. Die Strukturen sind dadurch bedingt, daß die Wellenfunktion des Endzustandes die Überlagerung einer vom absorbierenden Atom auslaufenden Welle mit von den Nachbaratomen reflektierten Anteilen darstellt. Beim Abstand R eines Nachbarn haben die reflektierten Wellen am Ausgangsatom eine Phasenverschiebung $2kR + \delta$, wobei δ durch die Streueigenschaften der atomaren Potentiale bestimmt ist. Eine Fourier-Analyse der EXAFS-Daten führt auf eine genaue Lagebestimmung der Nachbaratome. Damit ist die Strukturbestimmung von amorphen Festkörpern und Komplexverbindungen möglich.

Compton-Effekt. Neben der elastischen Rayleigh-Streuung der Röntgen-Photonen ist deren unelastische Streuung an kernnahen gebundenen Elektronen möglich. Ein Beispiel dafür ist der Röntgen-Fotoeffekt, bei dem die Gesamtenergie des Photons auf das emittierte Fotoelektron übertragen wird. Die Wechselwirkung kann aber auch nur zu einer Impulsänderung und zu einem partiellen Energieverlust des gestreuten Photons führen. Ebenso verändert das streuende Elektron im Festkörper Impuls und Energie (Abb.

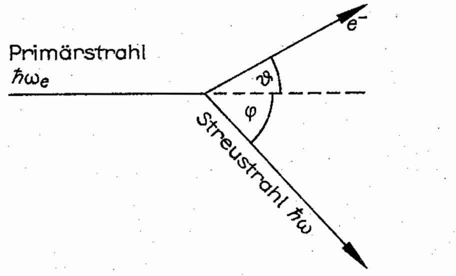


Abb. 6.8.1–25 Compton-Effekt am ruhenden Elektron

6.8.1–25). Energie- und Impulssatz geben darüber Auskunft: $\hbar\omega_e = \hbar\omega + mv^2/2$; $\hbar\omega_e/c = (\hbar\omega/c)\cos\varphi + mv\cos\vartheta$ (Impulskomponente parallel zum Primärstrahl); $(\hbar\omega/c)\sin\varphi + mv\sin\vartheta = 0$ (Impulskomponente senkrecht zum Primärstrahl). Die Wellenlängenänderung für die Compton-Streustrahlung ist $\Delta\lambda = \lambda_C 2\sin^2(\varphi/2)$ mit der Compton-Wellenlänge $\lambda_C = h/(mc)$.

Die Wellenlängenverschiebung der Compton-Linie ändert sich, wenn das streuende Elektron eine von Null verschiedene Geschwindigkeit hat (Leitungsbandelektron). Das Profil der Compton-Streuung informiert über die Geschwindigkeits- bzw. Impulsverteilung $n(p)$ der Elektronen im Festkörper. Im Experiment dazu trifft ein scharf begrenzter Photonenstrahl den Festkörper. Gemessen wird die Intensität der Streustrahlung in Abhängigkeit vom Streuwinkel φ. Die Energieverteilung der Compton-Streuung $I(\Delta\lambda)$ ist proportional einer linearen Impulsverteilung $I(p_z) = \int\int n(p)\mathrm{d}p_x\mathrm{d}p_y$ (z = Streurichtung des Photons). Für einfache Metalle ergeben sich nahezu parabelförmige Compton-Profile (Abb. 6.8.1–26). Die Abweichungen von der Parabelform sind charakteristisch für die Bandstruktur und die Bloch-Funktionen der Elektronen sowie für die Elektronenkorrelation. Vorteilhaft ist es, die Compton-Spektren mit harter γ-Strahlung zu messen. Die hohe Monochromasie verbessert u. a. die Auflösung der Spektren, die hohe Strahlungsenergie garantiert kurze Meßzeiten. Fotoeffekt und Compton-Effekt sind konkurrierende Prozesse. Ihr Wirkungsquerschnitt ist von der Photonenenergie abhängig (Abb. 6.8.1–27).

Die Wahrscheinlichkeit für den Compton-Effekt wächst linear mit der Ordnungszahl Z des bestrahlten Festkörpers, für den Fotoeffekt hingegen mit Z^5.

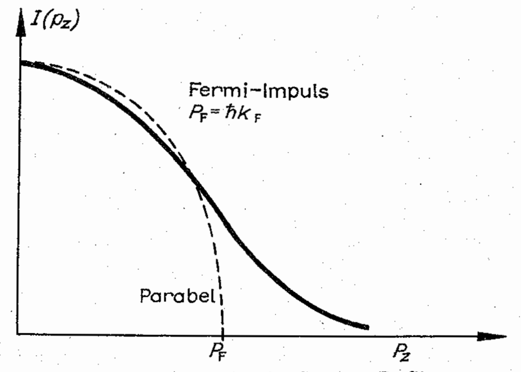


Abb. 6.8.1–26 Eindimensionales Compton-Profil

Röntgen-Fotoeffekt. Die durch Strahlungsabsorption emittierten Fotoelektronen können sowohl Valenz- als auch Rumpfelektronen des Festkörpers sein. Entsprechend ihrer unterschiedlichen Bindungsenergie haben sie nach der Emission verschiedene kinetische Energien. Deren Registrierung ist Gegenstand der *Röntgen-Fotoelektronenspektroskopie* (XPS). Der Emissionsvorgang verläuft als Dreistufenprozeß (vgl. 6.8.1.7.). Die kinetische Energie der Fotoelektronen folgt aus dem Energiesatz $E_{kin} = \hbar\omega - E_B^F - \varphi_{Sp}$ (Abb. 6.8.1–28). Aufgrund der energiereicheren Röntgen-Strahlung haben die Fotoelektronen gegenüber denen der Ultraviolettanregung (vgl. 6.8.1.7.) eine größere mittlere Austrittstiefe λ (vgl. Abb. 6.8.1–15).

Die Energieverlustmechanismen sind die gleichen wie bei UPS: Anregung und Ionisierung von Atomen, Plasmonenanregung, diffuse Elektron-Phonon-Streuung. Die Intensität des Fotoelektronenstromes I_i ist abhängig vom angereg-

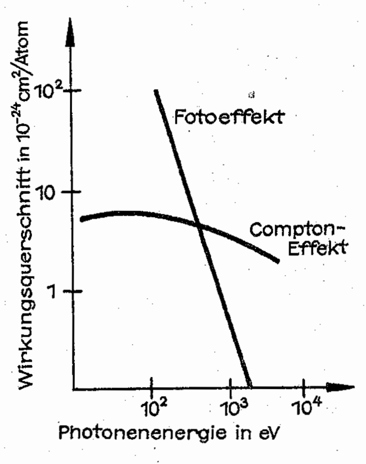


Abb. 6.8.1–27 Wirkungsquerschnitte für unelastische Wechselwirkungen zwischen Röntgen-Strahlung und Festkörper

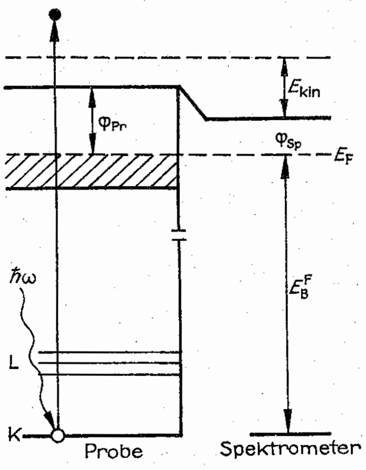


Abb. 6.8.1–28 Energiebandschema zum Röntgen-Fotoeffekt

ten Elektronenorbital i (z.B. $i = 2p_{3/2}$). Das Orbital wird auf verschiedene Weise wirksam, $I_i \sim \sigma_i \lambda_i N_i K_i$. Das geschieht über den fotoelektronischen Wirkungsquerschnitt σ_i, die mittlere freie Weglänge λ_i, die Anzahl der Atome N_i und einen Kontaminationsfaktor K_i (mit $K_i \leq 1$), der wirksam wird, wenn die Oberfläche von einer Kontaminationsschicht bedeckt ist. Die Nachweisempfindlichkeit von XPS als Analysenmethode beträgt etwa 10^{-3} µg/cm². Im Idealfall hat das Emissionsspektrum der Fotoelektronen aus Rumpfzuständen die Gestalt einer Lorentz-Kurve:
$I = I_{max}\Gamma_{1/2}^2/[4(E - E_{max})^2 + (\Gamma_{1/2}/2)^2]$. Durch die instrumentelle Verbreiterung und durch diffuse Streuung an Phononen wird daraus eine Gauß-Kurve:
$I = I_{max}\exp[-2(E - E_{max})^2/\Gamma_{1/2}^2]$ (Abb. 6.8.1–29). Eine Veränderung der chemischen Umgebung, d. h. der elektronischen Ladungsverteilung des betrachteten Atoms, hervorgerufen z. B. durch eine Oxidation des Festkörpers, führt zur Veränderung der Halbwertsbreite $\Gamma_{1/2}$ und der Peaklage E_{max}. Ladungsübergänge durch Verbindungsbildung sind deutlich erkennbar im Bereich der Valenzelektronen. Die Intensitätsverteilung im Fotoelektronenspektrum des Valenzbandes ist bestimmt durch den fotoelektrischen Wirkungsquerschnitt (σ) sowie die elektronischen Zustandsdichten von Ausgangs- bzw. Endzustand $N(E)$ bzw. $N(E + \hbar\omega)$ (ω = Frequenz der anregenden Röntgen-Strahlung.

Röntgen-Fluoreszenz. Ein in einem Rumpfniveau vorhandenes Elektronenloch (Abb. 6.8.1–30) kann durch ein Elektron aus einer anderen Elektronenschale wieder aufgefüllt werden. Dabei kommt es zur Emission von Röntgen-Fluoreszenzstrahlung. Die Strahlung wird durch ein starkes Röntgen-Bremskontinuum angeregt.

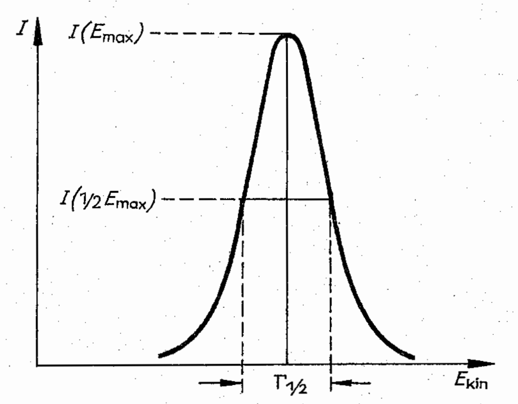


Abb. 6.8.1–29 XPS-Spektrum eines Rumpfelektronenniveaus

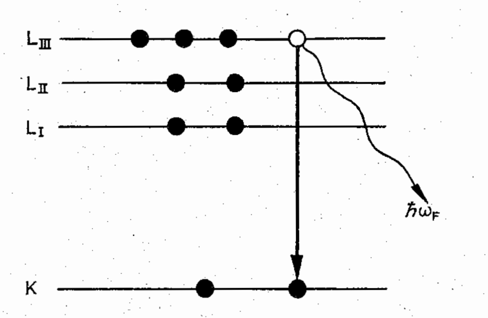


Abb. 6.8.1–30 Entstehung von Röntgen-Fluoreszenzstrahlung

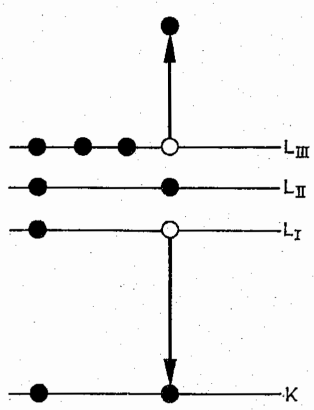


Abb. 6.8.1–31 Entstehung des Auger-Spektrums vom Typ KL_IL_{III}

Die Eigenstrahlung ist von der Art der anregenden Röntgenstrahlung abhängig. Sie bildet ein Linienspektrum, das sich aus mehreren Serien zusammensetzt, die bei entsprechend großer Primärenergie angeregt werden können. Die K-Serie kann nur angeregt werden, wenn die Primärenergie größer ist als die der K-Absorptionskante. Das emittierte Linienspektrum ist charakteristisch für die Atomart. In einem Röntgen-Spektrometer kann die nach Wellenlängen oder nach der Frequenz „zerlegte" Fluoreszenzstrahlung zur qualitativen und quantitativen Festkörperanalyse benutzt werden. Die Intensität der Fluoreszenzstrahlung (Fluoreszenzausbeute) wächst mit steigender Ordnungszahl Z. Sie beträgt für das K-Spektrum des Elementes Germanium (Z = 32) gerade 50 %. Bei dieser *Röntgen-Fluoreszenzanalyse* (RFA) werden Schichtdicken von 1 mm durchstrahlt, der Röntgen-Brennfleck (Fokus) beträgt 0,1 mm. Die Analyse, deren Nachweisempfindlichkeit 10^{-4} % beträgt, ist als Vielelementebestimmung sehr schnell durchführbar. Innerhalb von 10 min wird man über den Gehalt von sechs Elementen in einer Probe informiert. Problematisch wegen der geringen Fluoreszenzausbeute ist die Bestimmung leichter Elemente.

Auger-Elektronen. In Konkurrenz zur Fluoreszenzanregung steht ein strahlungsloser Elektronenübergang, in dessen Ergebnis anstelle von Röntgen-Quanten Elektronen (Auger-Elektronen) aus dem Festkörper emittiert werden. An dieser unelastischen Strahlungswechselwirkung sind i. allg. drei Elektronenenergieniveaus beteiligt (Abb. 6.8.1–31). Die kinetische Energie des emittierten Elektrons für den Übergang ABC ist: $E_{ABC} = E_A(Z) - E_B(Z) - E_C(Z + W)$, wobei die Bindungsenergien der Subschalen A, B, C im neutralen Atom (Ordnungszahl Z) maßgeblich sind. Für das C-Orbital ist nicht die Bindungsenergie des neutralen, sondern des einfach ionisierten Festkörperatoms entscheidend. Die damit verbundene Änderung der effektiven Kernladungszahl wird durch die Korrekturgröße $W(1/2 \leq W \leq 3/4)$ berücksichtigt. Neben der Besonderheit des einfach ionisierten Zustandes ist auch der doppelt ionisierte Endzustand in einer exakten Energiebilanz zu berücksichtigen. Er äußert sich in elektrostatischen Wechselwirkungskräften zwischen den Elektronenwolken der nicht gefüllten Schalen und in magnetischen Wechselwirkungen. Besondere Auger-Übergänge sind die Coster-Kronig-(AAC) bzw. Super-Coster-

Kronig-Übergänge (AAA). Wenn das Niveau A durch die K-Schale repräsentiert wird, ist für den Ordnungszahlbereich $Z > 32$ die Anregungswahrscheinlichkeit des Auger-Überganges wesentlich kleiner als die der Fluoreszenzanregung (Abb. 6.8.1–32).

Außerdem werden mit steigender Ordnungszahl die Lebensdauern der angeregten Lochzustände immer kleiner, so daß die Spektrallinien des Auger-Spektrums immer breiter werden und im statistischen Untergrund verschwinden. Bei höherenergetischen A-Energien verschiebt sich der Kreuzungspunkt der Wahrscheinlichkeiten nach höheren Z-Werten. Bei Röntgen-Anregung gibt es im Z-Bereich 1–12 deutliche Auger-Übergänge des Typs KBC, im Bereich 13–33 des Typs LBC. Natürlich können Auger-Elektronen auch durch Elektronenbeschuß des Festkörpers erzeugt werden (vgl. 6.8.2.1.).

Mössbauer-Effekt. Die von Mössbauer-Atomen, welche in zu untersuchenden Kristallgittern eingelagert sind, resonanzabsorbierte γ-Strahlung (vgl. 5.11.) zeigt indirekt Wechselwirkungen mit dem Wirtsgitter des Absorbermateriales. Verändern sich die Eigenschaften der Elektronenhülle (z. B. die räumliche Verteilung der Ladungsdichte) der Atome des Wirtsgitters, so werden durch die Hyperfeinwechselwirkung auch die Energieniveaus der Mössbauer-Atome, die auf energetisch verschiedenen Gitterplätzen angeordnet sein können, beeinflußt. Diese Reaktion ist gitterplatzabhängig in der chemischen Verschiebung (Isomerieverschiebung) der Mössbauer-Linie ablesbar.

Paarerzeugung. Für energiereiche Photonen des γ-Bereiches besteht beim Beschuß eines Festkörpers die Möglichkeit der Paarerzeugung. Die Photonenenergie $\hbar\omega$ muß groß gegen 1 MeV sein. Das Quant verschwindet, und es bildet sich durch die Absorption ein Elektron-Positron Paar. Die Energiebilanzgleichung lautet $E_{\text{kin}}^{e^-} + E_{\text{kin}}^{e^+} = \hbar\omega - 2mc^2$ mit mc^2 als Ruheenergie des freien Elektrons (511 keV). Das Positron e^+ zerstrahlt wieder mit einem Elektron in zwei γ-Quanten der Energie 511 keV.

6.8.2. Wechselwirkung mit geladenen Teilchen

6.8.2.1. Strahlungsdualismus. 1924 übertrug de Broglie den Dualismus der elektromagnetischen Strahlung (vgl. 6.8.1.) auf Materie. Danach haben bewegte Teilchen (Elektronen, Neutronen) Welleneigenschaften. Einem Strom von Teilchen mit dem Impuls p ist eine Wellenlänge $\lambda = h/p$ (*de Broglie-Wellenlänge*) zuzuordnen. Aus $E = c\sqrt{(mc)^2 + p^2}$ folgt $v_{gr} v_{ph} = c^2$ für das Produkt von Gruppen- und Phasengeschwindigkeit. Wellen- und Teilcheneigenschaften im dualen Erscheinungsbild sind durch die Heisenbergsche Unschärferelation miteinander verbunden.

6.8.2.2. Wechselwirkung mit Elektronenstrahlen. Wird ein Festkörper mit Elektronen beschossen, so sind die Wechselwirkungsphänomene abhängig von den Eigenschaften des Elektronenstrahls und des Festkörpers. Die Primärelektronen können elastisch oder unelastisch vom Festkörper reflektiert, absorbiert werden bzw. denselben durchdringen. Außerdem können *Sekundärelektronen*, das sind festkörpereigene Elektronen, durch Energieaustausch emittiert werden (Abb. 6.8.2–1). Absorption und Transmission der Primärelektronen bewirken die Anregung von Plasmaschwingungen der Leitungselektronen und die Bildung von Exzitonen. Es kann zur Anregung von Gitterschwingungen, zu atomaren Umlagerungen und zur Aussendung von Röntgen- sowie Lumineszenzstrahlung kommen.

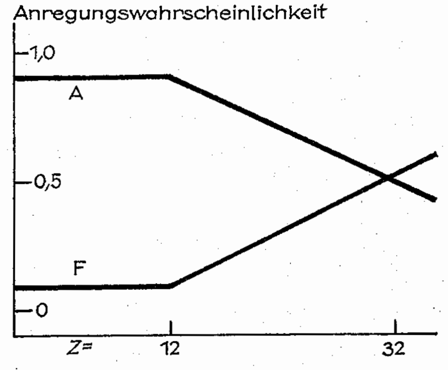


Abb. 6.8.1–32 Anregungswahrscheinlichkeit in Abhängigkeit von der Ordnungszahl Z

Aufgrund der Welleneigenschaften der Materiewellen treten im Primärstrahl Beugungs- und Interferenzerscheinungen auf.

Elektronenbeugung. Die erste experimentelle Bestätigung der Ideen von de Broglie erfolgte 1927 durch Davisson und Germer sowie durch Thomson und Reid. Sie erzeugten mit Elektronenstrahlen an Kristallen Beugungseffekte, die Resultate analog zur Röntgen-Strahlbeugung (vgl. 6.8.1.8.) darstellten. Bei bekannten Atomabständen im Kristallgitter läßt sich aus der Anordnung der Beugungsreflexe die Wellenlänge der Elektronenstrahlung bestimmen. Sie entspricht dem Wert von de Broglie. Mit diesem können Gitterkonstanten berechnet werden. Die Beschreibung der Elektronenstrahlbeugung am Festkörper erfolgt mit Hilfe eines ähnlichen Formalismus wie er von der Röntgen-Strahlbeugung bekannt ist (Bragg-Bedingung, Ewald-Konstruktion, dynamische Theorie, vgl. 6.8.1.8.). Dennoch gibt es einige wesentliche Unterschiede.

Während bei Röntgen-Wellen (Kugelwellen) die dynamische Theorie nur zur Erklärung von Feinheiten des Beugungsbildes herangezogen werden muß, dominiert bei den Elektronenwellen der Einfluß der Vielfachstreuung, die Elektronenwellen haben den Charakter von Bloch-Wellen. Die Wechselwirkung zwischen Strahlung und Festkörper im

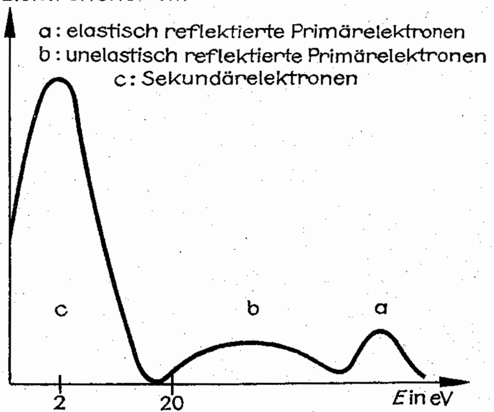


Abb. 6.8.2–1 Elektronenspektrum

Röntgen-Strahlbereich kann bei Wellenlängen zwischen 0,23 nm (Chromiumstrahlung, $Z = 24$) und 0,015 nm (Goldstrahlung, $Z = 79$) verfolgt werden. Die Variation der Elektronenbeschleunigungsspannung von einigen Volt bis zu ≈ 3 MV bietet dagegen ein Wellenlängenspektrum von $\approx 0{,}1$ nm bis zu $5 \cdot 10^{-4}$ nm an. Entsprechend differenziert sind die Wechselwirkungsphänomene, so daß eine Einteilung in zwei Beugungsgebiete zweckmäßig ist.

Bei der Beugung langsamer Elektronen (LEED: *L*ow *E*nergy *E*lectron *D*iffraction) beträgt die Beschleunigungsspannung maximal 2 kV. Die Wellenlängen sind in der Größenordnung der Gitterkonstanten der Kristalle. Beruhte die Schwächung von Röntgenstrahlung im Festkörper auf einer Wechselwirkung mit der elektronischen Ladungsdichte (Elektronenhülle) der Atome, so ist für die Elektronenstrahlschwächung ein anderer Mechanismus wirksam. Elastische und unelastische Streuung erfolgen nicht nur an der Elektronenhülle, sondern auch am Atomkern, wodurch eine $\approx 10^8$fache Erhöhung des Wirkungsquerschnittes zu verzeichnen ist. Diese Überlegungen gelten für einen einfachen Streuakt. In Wirklichkeit geschieht der Impuls- und Energieaustausch über eine Vielfachstreuung. Für LEED heißt das, daß die Erfüllung der Braggschen Interferenzbedingung nicht auf einen Punkt im reziproken Gitter beschränkt ist, wie bei der Zweistrahlinterferenz im Falle der Röntgen-Strahlbeugung (vgl. 6.8.1.8.), sondern für ein ausgedehntes Gebiet. Das hängt mit der wegen des großen Streuquerschnittes geringen Eindringtiefe der Elektronen in den Festkörper zusammen. Sie beträgt einige 10^{-1} nm (bzw. Monolagen). Deshalb „sehen" die Elektronen in der Einschußrichtung anstelle des Punktes im reziproken Gitter einen Stab (entarteter Gitterpunkt). Die atomistische Struktur senkrecht zur Oberfläche „geht verloren". Dagegen reagieren die LEED-Beugungsbilder empfindlich auf Veränderungen des Kristallgitters an der Festkörperoberfläche. Die Gittersymmetrie an der Oberfläche spiegelt sich wider in der Symmetrie des Beugungsbildes, das ein vergrößertes Bild des reziproken Gitters des Kristalles darstellt. Auch die Intensitätsverteilung der Beugungsreflexe enthält Informationen über die Struktur der Festkörperoberfläche. Allerdings ist eine Auswertung schwieriger als bei der Röntgen-Strukturanalyse wegen des komplizierteren Streumechanismus der Elektronen.

Durch Erhöhung der Beschleunigungsspannung der Elektronen wird deren Eindringvermögen in den Festkörper erhöht. Durchlässigkeit wird bei entsprechend dünn präparierten Folien (Dicke im Bereich von µm) erreicht. Die Beobachtung von HEED (*H*igh *E*nergy *E*lectron *D*iffraction) ist somit in Reflexion (RHEED) und in Transmission (THEED) möglich. Die Ausdehnung des Interferenzbereiches gegenüber LEED nimmt bei HEED stark ab. Da aber der Krümmungsradius der Ewaldschen Ausbreitungskugel wächst und somit Reflexionsbereiche von vielen Punkten des reziproken Gitters tangiert, bleibt der Effekt der Vielfachstreuung erhalten. Das gilt um so mehr, je höher die Beschleunigungsspannung der Elektronen ist. Das Unschärfegebiet für die Erfüllung der Bragg-Bedingung liegt in der Größenordnung des Bragg-Winkels, wogegen es für die Röntgen-Strahlbeugung $\approx 10^{-3}$ mal kleiner ist. Auch HEED informiert mit der Abbildung des reziproken Kristallgitters über den Gitteraufbau und die Kristallorientierung. Bei einkristallinen Proben bestehen die Beugungsdiagramme aus Punktreflexen. Zusätzliche orientierungsabhängige Beugungslinien, die *Kikuchi-Linien*, treten bei unelastischer Wechselwirkung der Elektronen im Festkörper auf.

Mit Hilfe der *Elektronenmikroskopie* gelingt die direkte Abbildung mikroskopischer Festkörperbereiche. Unter Verwendung spezieller Abbildungsverfahren (Elektronenlinsen, Elektronenspiegel, Rastertechnik) werden Objektvergrößerungen von 10^6 erreicht. Die elektronenmikroskopischen Bilder informieren über die Morphologie des Festkörpers und ergänzen die durch Beugungseffekte bestimmten Strukturdaten (vgl. 7.7.).

Elektronenstrahlmikroanalyse (ESMA). Der analytischen Untersuchung von Festkörpern in Mikrobereichen dient die Spektroskopie der primären Röntgen-Strahlung, die nach Beschuß des Festkörpers mit einem eng gebündelten Elektronenstrahl entsteht. Das analysierte Probenvolumen ist abhängig von der Eindringtiefe der Elektronen r_e und von der Anregungstiefe r_x der Röntgen-Strahlung. Es hat die Größenordnung von einigen Kubikmikrometern. Die Parameter r_e und r_x sind von denselben Größen abhängig: von der Primärenergie der Elektronen E_0, von der Ionisierungsenergie E_i der Rumpfelektronenniveaus des Festkörpers, der Ordnungszahl Z und dem Atomgewicht A: $r_e = (1{,}54\, A/\varrho Z^{0{,}8})(E_0^{1{,}8} - E_i^{1{,}8})$ sowie $r_x = (3{,}3\, A/\varrho Z)(E_0^{1{,}7} - E_i^{1{,}7})$ (E_0, E_i in keV, r_e, r_x in 10^{-4} m). Die spektrale Zerlegung des emittierten Röntgen-Spektrums kann mit Hilfe von Spektrometerkristallen nach der Braggschen Gleichung wellenlängendispersiv erfolgen. Ein anderes Spektrometer benutzt die Energie der Röntgen-Strahlung zur Erzeugung von Ladungsträgern in Festkörperdetektoren auf Halbleiterbasis (Ge(Li)-Detektoren). Die Konzentration der freigesetzten Ladungsträger wird energiedispersiv und simultan in elektronischen Zählgeräten registriert und analysiert. Die Nachweisempfindlichkeit im mittleren Ordnungszahlbereich ist für die wellenlängendispersive Nachweismethode etwa um eine Größenordnung höher als für die energiedispersive. Deren Vorteil besteht in einer wesentlich kürzeren Analysendauer. Zu ihren Nachteilen rechnet auch, daß ein Nachweis leichter Elemente im Festkörper noch schwieriger ist (für $Z \leq 13$) als beim Kristallspektrometer. Zur *quantitativen Analyse* müssen die Faktoren, die die Intensität der einzelnen Röntgen-Spektrallinien beeinflussen, bekannt sein. Dazu gehören neben der Massendicke des jeweiligen Elementes der Elektronenrückstreufaktor, unelastische Elektronenenergieverluste, der Ionisierungsquerschnitt und die Sekundärstrahlenausbeute (Fluoreszenzstrahlung).

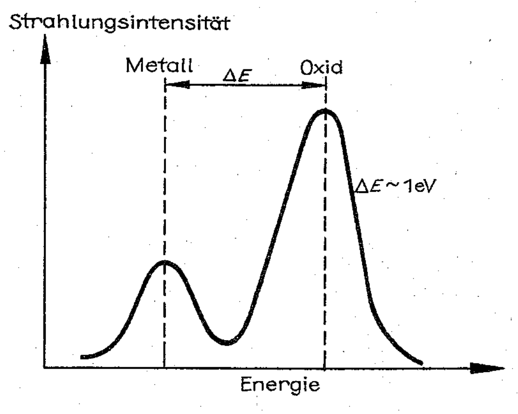


Abb. 6.8.2-2 Linienverschiebung im Röntgen-Spektrum bei Oxidation des Festkörpers

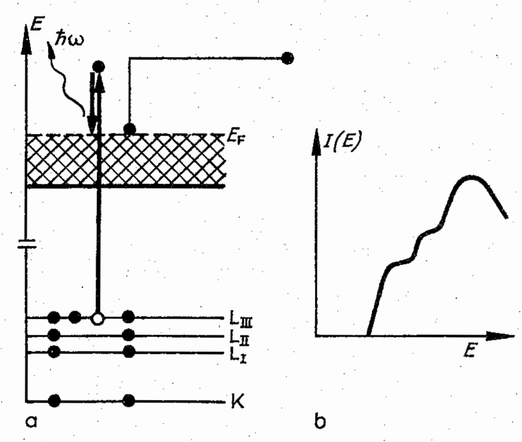


Abb. 6.8.2–3 (a) Isochromatenspektroskopie, (b) Isochromatenspektrum

Es ist bequemer, sich Probenstandards herzustellen und die entsprechenden Intensitäten der Röntgen-Strahlen vom Meßobjekt auf die *Standardintensitäten* zu beziehen. In das Verhältnis der Intensitäten geht dann lediglich die Konzentration des zu bestimmenden Elementes ein. Das ist eine stark vereinfachte Überlegung, die davon ausgeht, daß die mittlere Ordnungszahl von Probe und Standard übereinstimmen. Wenn das nicht der Fall ist, muß eine Analysenkorrektur vorgenommen werden, die die Ordnungszahldifferenz berücksichtigt. Weiterhin haben die Strahlungsabsorption und die Fluoreszenz einen Einfluß auf das Intensitätsverhältnis. Bei Röntgen-Linien, die durch Elektronenübergänge in äußeren Schalen emittiert werden, wird die Strahlungsenergie vom Zustand der chemischen Bindung beeinflußt. Die Oxidation von Metallen oder Halbleitern führt zur *chemischen Verschiebung* und zur Änderung der energieabhängigen Intensitätsverteilung der Röntgen-Spektrallinie (Abb. 6.8.2–2).

Isochromatenspektroskopie. Schießt man Elektronen einer bestimmten Energie auf den Festkörper und registriert die erzeugte Röntgenstrahlung bei einer bestimmten festgehaltenen Frequenz (Isochromaten = Linien „gleicher Farbe") in Abhängigkeit von der variierbaren Elektronenenergie, so erhält man analog zur Röntgen-Absorptionsspektroskopie Informationen über die unbesetzten Elektronenzustände des Festkörpers (Abb. 6.8.2–3).

Die Isochromatenspektroskopie ist die inverse XPS (vgl. 6.8.1.8.). Sie ergänzt diese vorzüglich, indem jetzt die unbesetzten Zustände des Valenz- (oder Leitungs-)bandes abgebildet werden.

Röntgenstrahl-Appearancespektroskopie (Schwellenwert-Anregungsspektroskopie). Wird die Beschleunigungsspannung der auf einen Festkörper geschossenen Elektronen erhöht, so nimmt bei einer ganz bestimmten Spannung die Ausbeute der emittierten Röntgen-Strahlung ab. Diese Elektronenenergie ist charakteristisch für die Anregung eines Rumpfelektronenniveaus. Das freigesetzte Elektron geht in unbesetzte Zustände oberhalb des Fermi-Niveaus E_F über (Abb. 6.8.2–4a). In den Appearancespektren ist die Abhängigkeit der ersten Ableitung der totalen emittierten Röntgen-Strahlintensität dI/dE über der Beschleunigungsspannung der Elektronen aufgetragen. Die Spektren sind stark strukturiert (Abb. 6.8.2–4b).

Das erste Maximum ist der Wendepunkt der Absorptionskante und informiert über die von den Rumpfelektronen aufgenommene Energie. Ergeben sich aufgrund chemischer Änderungen energetische Verschiebungen der Rumpfelektronenniveaus, so sind diese als chemische Verschiebungen in den Spektren nachweisbar. Die komplizierte Spektrenstruktur ist ein Ausdruck dafür, daß neben der Anregung eines Elektronen-Loch-Paares noch Plasmonen angeregt werden. Beide Anregungszustände können getrennt werden, indem mit einem gut kollimierten Elektronenstrahl die Messungen an einkristallinen Oberflächen mit veränderlichem Einfallswinkel der Primärelektronen durchgeführt werden.

Sekundärelektronen. Neben der Erzeugung primärer charakteristischer Röntgen-Strahlung durch die Elektronensonde gibt es weitere Wechselwirkungsphänomene mit dem Festkörper. Sie haben unterschiedliche relative Wirkungstiefen. Die Absolutwerte hängen ab von der Primärenergie der Elektronen, der Anregungsenergie des angeregten Zustandes und von der Massendichte. Die Größenordnung der maximalen Informationstiefe des Kaskadenprozesses (Abb. 6.8.2–5) liegt im Bereich von einigen µm.

Die *Sekundärelektronenausbeute* ist unabhängig von der Ordnungszahl des Elementes. Empfindlich reagiert die Aus-

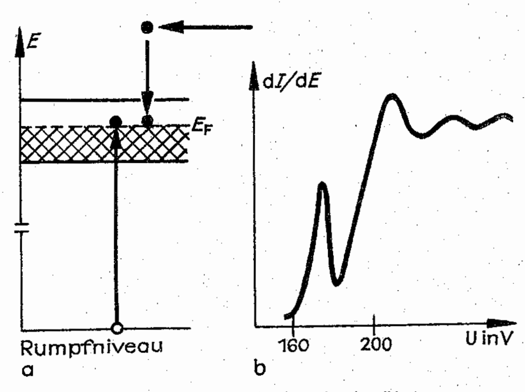


Abb. 6.8.2–4 (a) Appearancespektroskopie, (b) Appearancespektrum

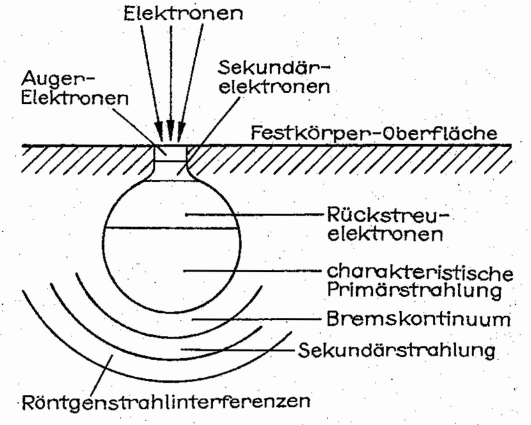


Abb. 6.8.2–5 Wechselwirkungskaskade von Elektronen im Festkörper

beute auf die Beschaffenheit der Oberfläche (Kontamination, elektrische und magnetische Inhomogenitäten, Rauhigkeit). Dagegen besteht keine deutliche Korrelation zur Austrittsarbeit an sauberen einkristallinen Oberflächen. Durch die Veränderung des Einfallswinkels der Primärelektronen wird über die Ionisierungsstrecke die Sekundärelektronenkonzentration moduliert. Besondere Bedeutung haben die Sekundärelektronen in der Rasterelektronenmikroskopie (vgl. 7.7.). Rauhe Oberflächen geben im Sekundärelektronenbild einen guten Kontrast. Unter bestimmten Bedingungen treten Beugungseffekte in Form von orientierungsabhängigen *Pseudo-Kikuchi-Linien* auf.

Katodolumineszenz. Die Rekombination elektronenstrahlangeregter Elektronen-Loch-Paare ergibt eine Lumineszenzstrahlung im Wellenlängenbereich von einigen Mikrometern bis zu einigen 10^{-1} µm. Die Strahlung kann nach dem Mikrosondenprinzip mit Hilfe von Sekundärelektronenvervielfachern registriert werden. Besonderen Einfluß auf die Katodolumineszenz haben Kristallbaufehler, die als Rekombinationszentren wirken.

Elektronen-Energieverlust-Spektroskopie (EELS). Mit Hilfe eines Elektronenspektrometers können unelastisch gestreute Primärelektronen zur Eigenschaftsbestimmung des Streumediums analysiert werden. Die Energieverluste sind abhängig von der Primärelektronenenergie, vom Streuwinkel, von der Elektronenkonzentration des Festkörpers im Wechselwirkungsvolumen sowie von der Energieverlustfunktion $-\mathrm{Im}[\varepsilon^{-1}(\omega)]$ (vgl. 6.8.1.6.). Der Verlustmechanismus ist die Ionisierung innerer Elektronenniveaus des Festkörpers. Die optischen Auswahlregeln haben bei Elektronen-Atom-Stößen keine Gültigkeit, so daß mit der EELS optisch verbotene Übergänge (Singulett-Triplett- und Quadrupol-Übergänge) untersucht werden können. Hauptanwendungsgebiet der Verlustspektroskopie ist die Untersuchung von Plasmonenspektren im Volumen und an der Oberfläche kondensierter Materie.

Elektronenkanalleitung. Eine besondere Form der Wechselwirkung zwischen schnellen Elektronen und fester Materie tritt auf, wenn die Elektronen in „Kanäle" eines Einkristalles, d.h. in seinen Zwischengitterbereich hineingeschossen werden (Abb. 6.8.2–6).

Beim „richtigen" Einschußwinkel dringen die Elektronen wesentlich tiefer in den Festkörper ein. Sind die Kanäle durch Fremdatome (Verunreinigungen) verstopft, so liefern rückgestreute Kanalelektronen Informationen über die „Tiefe" des besetzten Zwischengitterplatzes. Beim Einsatz von Protonen anstelle der Elektronen kann auch die Masse des Zwischengitteratoms ermittelt werden.

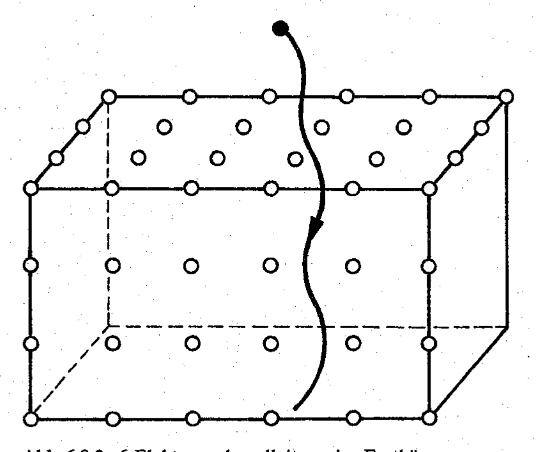


Abb. 6.8.2–6 Elektronenkanalleitung im Festkörper

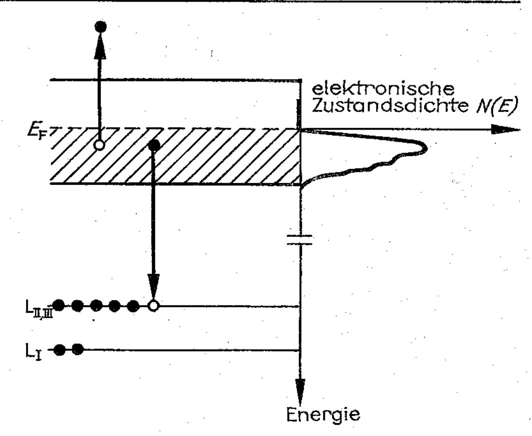


Abb. 6.8.2–7 Auger-Übergang vom Typ LVV

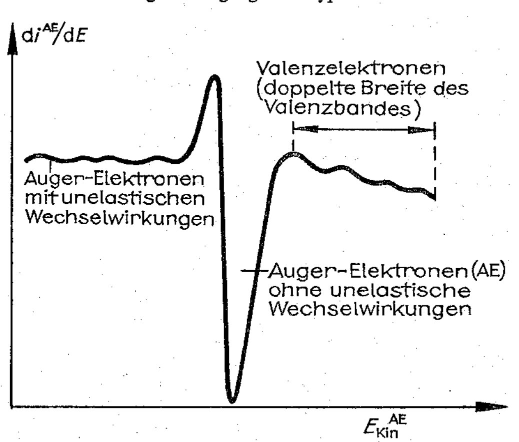


Abb. 6.8.2–8 Differenziertes AE-Spektrum

Auger-Elektronenspektroskopie (AES). Neben Röntgen-Strahlanregung (vgl. 6.8.1.8.) können Auger-Elektronen auch durch schnelle Elektronen erzeugt werden. Die Energieverteilung der Auger-Elektronen ist festkörperspezifisch. Eine teilweise Bedeckung der ersten atomaren Lage des Festkörpers mit Verunreinigungen (adsorbierte Gase) von der Größenordnung 10^{-3} Monolagen (d.h. 10^{12} Atome/cm²) kann noch nachgewiesen werden. Am Auger-Übergang können auch Valenzelektronen beteiligt sein (Abb. 6.8.2–7).

Der Übergang AVV bildet die energieabhängige elektronische Zustandsdichte im Valenzband eines Festkörpers ab. Auch bei Elektronenanregung gibt es die Konkurrenz zwischen der Elektronenanregung und der Fluoreszenzstrahlung (vgl. 6.8.1.8.). Die Auger-Elektronenemissionsrate ist hauptsächlich festgelegt durch die Ionisierungsrate des A-Niveaus. Für eine Bindungsenergie von $E_A < 500$ eV dominiert die Emissionsrate, der Fluoreszenzanteil der Energieumwandlung ist vernachlässigbar. Bei $E_A \approx 2$ keV sind Auger-Effizienz und Sekundärstrahlungsrate etwa gleich groß. Der Ordnungszahleinfluß ist gering. Auf chemische

Veränderungen im Festkörper reagiert das Auger-Elektronenspektrum. Die integrale chemische Verschiebung folgt (in starker Vereinfachung) aus der Verschiebung der Einzelniveaus: $\Delta E = -\Delta E_A + \Delta E_B + \Delta E_C$. Eine genauere Energieanalyse mittels AES ist möglich, wenn an Stelle des Elektronenspektrums das energiedifferenzierte Verteilungsspektrum (1. Ableitung) aufgenommen wird. Der Auger-Übergang ist in diesem Spektrum gut getrennt von Elektronen, deren kinetische Energie durch unelastische Wechselwirkung (Plasmonenanregung, Interbandübergänge, Elektronen-Loch-Paarbildung) geändert wurde (Abb. 6.8.2–8). Der Einsatz der AES für eine quantitative Analyse ist schwierig, da man die Austrittstiefe der Auger-Elektronen und die Tiefenverteilung der die Verbindung bzw. den Mischkristall bildenden Elemente kennen müßte. Deshalb hat sich die Verwendung von Standardproben bewährt. AES und LEED bilden eine ideale Kombination zur Untersuchung von Festkörperoberflächen, erstere wegen ihrer hohen qualitativen Nachweisempfindlichkeit, letztere infolge ihrer Strukturempfindlichkeit.

6.8.2.3. Wechselwirkung mit Positronen. Positronen haben die gleiche Masse wie Elektronen, ihre Ladung, betragsmäßig gleich der der Elektronen, hat ein positives Vorzeichen, Positronen werden von künstlichen Radionukliden (Na^{22}, Co^{58}, Cu^{64}) emittiert. Elektronen des Festkörpers üben durch die Coulomb-Wechselwirkung eine starke Anziehungskraft auf Positronen aus. Positronen haben nach ihrer Emission aus dem Radionuklid eine hohe kinetische Energie (Größenordnung 10^6 eV). Beim Eindringen in den Festkörper verlieren sie innerhalb von 10^{-12} s durch Stoßprozesse mit den Festkörperatomen ihre Energie. Sie sinkt auf die der thermischen Energie des Festkörpers ($\approx 0{,}03$ eV bei Raumtemperatur). Dieser Vorgang heißt *Thermalisation.* Nach weiteren 10^{-9} s vereinigen sich die thermalisierten Positronen mit Valenz- bzw. Rumpfelektronen, wobei das Ladungspaar verschwindet, auf Kosten von emittierten γ-Quanten. Die Vernichtung der Positronen heißt *Positronenannihilation.* Die emittierte γ-Strahlung kann aus zwei oder drei Quanten je Teilchenpaar bestehen. Die Drei-Quantenstrahlung ist in kondensierter Materie 372mal unwahrscheinlicher und setzt bei einer höheren Intensität die Bildung von *Positronium,* einem dem Wasserstoffatom ähnlichen metastabilen Zustand, voraus. Wegen des zu kleinen Platzangebots kann sie in anorganischen Festkörpern i. allg. vernachlässigt werden.

Die *Positronenlebensdauer* sowie die Energie- bzw. Winkelverteilung der 2γ-Annihilationsstrahlung reagieren auf die Elektronendichteverteilung im Festkörper sowie auf Störungen des Kristallgitters durch Baufehler (Leerstellen, Versetzungen). Für ein ruhendes Elektron-Positron-Paar werden die beiden γ-Quanten mit einer Energie von jeweils $511\,\text{keV} = mc^2$ (m = Ruhemasse von Elektron bzw. Positron, c = Vakuumlichtgeschwindigkeit) unter einem Winkel von 180° emittiert. Hat das Ladungspaar einen Impuls $\boldsymbol{p} \neq 0$, so verkleinert sich der Emissionswinkel (Abb. 6.8.2–9) gemäß $\sin\Theta \approx \Theta = p_z/(m_e c)$. Die mit einem γ-Spektrometer gemessene Annihilationslinie bei 511 keV verbreitert sich um $\Delta E = p_y c/2$ *(Doppler-Verbreiterung).* Die *2γ-Winkelkorrelation* bestimmt die Θ-abhängige Intensität $I(\Theta)$ der Koinzidenzrate der γ-Quanten, die zum selben Annihilationsvorgang gehören. Die mit Szintillationsdetektoren registrierten Meßkurven setzen sich zusammen aus einer Parabel und aus einer Gauß-Kurve (Abb. 6.8.2–10).

Das Korrelationsdiagramm verändert sich durch Kristallbaufehler. Zwischengitterplatzbesetzung hat keinen Einfluß auf die Annihilation der Positronen. Verschiedene Atome in intermetallischen Verbindungen und Mischkristallen haben unterschiedliche Anteile an der Positronenannihilation. Aus der Koinzidenzrate der 2γ-Winkelkorrelation können Leerstellenkonzentrationen, Versetzungsdichten und Leerstellenbildungsenergien experimentell ermittelt werden.

6.8.2.4. Wechselwirkung mit Ionenstrahlen. Die Wechselwirkung zwischen Festkörperatomen und Ionenstrahlen hängt ab von der Energie E (eV- bis MeV-Bereich), der Ladung $Q_I e$ der eingeschossenen Ionen und den Eigenschaften des Festkörpertargets. Die Ionen führen elastische und unelastische Stöße mit den Elektronen der Targetatome aus sowie elastische und unelastische Stöße mit den Atomkernen. Kaum Bedeutung haben unelastische Kernstöße und elastische Elektronenstöße.

Unelastische Elektronenstreuung ändert Impuls und Energie der Ionen. Das stoßende Atom des Festkörpers wird je nach Ionenenergie angeregt oder ionisiert. Die Ionisation führt zur Elektronenemission, die Anregung zur Emission elektromagnetischer Strahlung.

Rutherford-Streuung. Der Atomkern der Festkörperatome ist von einem elektrischen Feld umgeben. Ein positives Ion (z. B. Ar^+, O^+) erfährt in diesem Feld eine abstoßende Coulomb-Kraft, wird aus dem Kernfeld weggestreut. Diese elastische Streuung ist die *Rutherford-Streuung* (*RBS, R*utherford *B*ack*s*cattering). Das Atom wird dabei nicht angeregt. Das Ion wird auf eine Hyperbelbahn gestreut

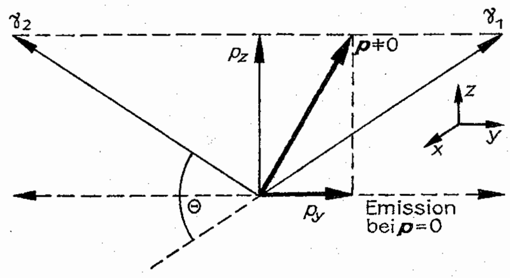


Abb. 6.8.2–9 Positronenannihilation

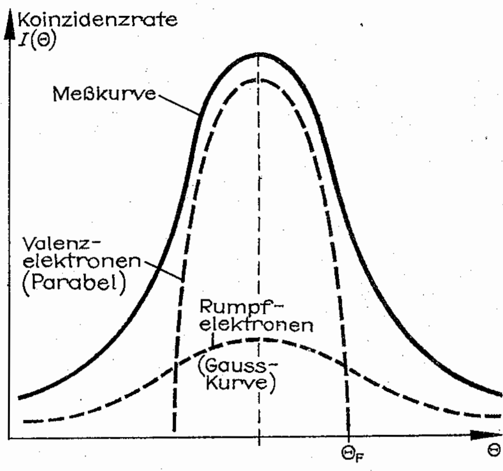


Abb. 6.8.2–10 2-γ-Winkelkorrelationsdiagramm

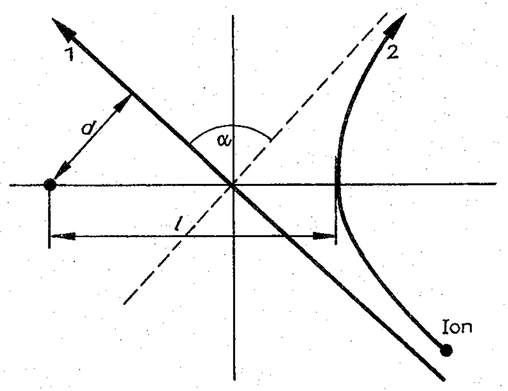


Abb. 6.8.2–11 Rutherford-Streuung im Zweiteilchenmodell

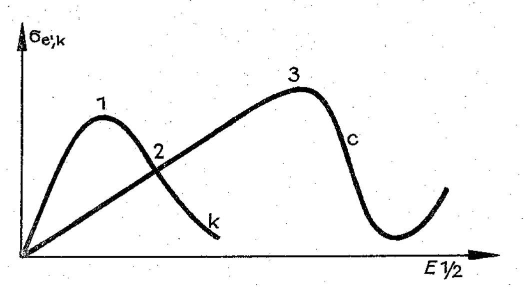


Abb. 6.8.2–12 Elektronischer und Kernbremsquerschnitt in Abhängigkeit von der kinetischen Energie der Ionen

(Abb. 6.8.2–11). Bei einer idealisierten *Zweiteilchenstreuung* sind die Stoßparameter $d = (l/2)\cot(\alpha/2)$.

Für größere Streuwinkel kommen die Ionen in Kernnähe, so daß neben der Coulomb-Streuung noch die direkte *Kernstreuung* wirksam werden kann. Für ein punktförmiges Ion ist bei einem elastischen Stoß der Wirkungsquerschnitt $\sigma_k = 2\pi R^2$ (R = Kernradius), weicht also von der Projektionsfläche des Kerns in Richtung der eingeschossenen Ionen (πR^2) ab. Das ist eine Folge der Beugungseffekte, die sich wegen des Wellencharakters der Teilchenströme (vgl. 6.8.2.1.) einstellen. Unter bestimmten Bedingungen können Targetkerne und Ionen kurzzeitig „Hyperkerne" mit gemeinsamer Elektronenhülle bilden. Die unelastische Wechselwirkung mit Elektronen und die elastische Kernwechselwirkung bilden einen integralen Ionenwechselwirkungseffekt. Ihre relativen Streuanteile bestimmt der jeweilige Streuquerschnitt $\sigma_{e;k} = -(1/n)(\mathrm{d}E/\mathrm{d}x)_{e;k}$ mit x als Ortskoordinate senkrecht zur Festkörperoberfläche und n als atomarer Dichte (Abb. 6.8.2–12). Die charakteristischen Werte 1–3 der Abb. 6.8.2–12 sind vom Ion, seiner Energie und vom Targetmaterial abhängig. Bei gleichem Target verschieben sich die Werte mit steigender Ionenenergie unterschiedlich nach höheren Energien.

Ionenimplantation. Die in einen Festkörper eindringenden Ionen durchlaufen eine Zick-Zack-Bahn und bleiben im Target stecken. Dieser Einbau von Fremdionen ist die *Ionenimplantation.* Die *Eindringtiefe* (Abb. 6.8.2.–13) der Ionen mit der Anfangsenergie E ist $R = \frac{1}{n}\int_0^E \mathrm{d}E/[\sigma_k(E) + \sigma_e(E)]$.

In Abhängigkeit von den Massen der Projektil-Ionen und der Targetatome sowie der Primärionenenergie bilden Festkörperatome Zwischengitterplatzatome oder lagern sich zu Atomgruppen *(Cluster)* zusammen. Bei genügend hoher Ionendosis wird der kristalline Körper amorphisiert.

Bei Raumtemperatur ist die kritische Dosis für leichtere Ionen $10^{16}/\mathrm{cm}^2$, für Schwerionen $\approx 10^{14}/\mathrm{cm}^2$. Bei einer Temperaturerhöhung steigt die kritische Dosis, da die Wärmebewegung (Diffusion) der Atomumlagerung entgegenwirkt.

Kanalleitung (Channeling-Effekt). Im kristallinen Material dringen etwa 25 % der Ionen tiefer in den Festkörper ein als im amorphen Zustand. Für regelmäßig angeordnete Atome (Einkristalle) gibt es Ionenbahnen mit sehr unterschiedlicher Reichweite. Sie sind abhängig vom Einschußort, vom Winkel der Ionen relativ zu den Netzebenen des Kristallgitters. Drei prinzipielle Ionenbahnen sind zu unterscheiden (Abb. 6.8.2–14). Tritt das Ion in der Mitte eines Gitterkanals ein (A), kann es sich im Kanal über eine große Strecke (μm-Bereich) nahezu kräftefrei bewegen. Entfernt sich der Einschußort von der Kanalmitte, ist ein Kanalwechsel möglich. Die Einzelbahnen bilden mit einer mittleren Bahnrichtung nur einen kleinen Winkel α (B). Im Fall (C) gibt es keine Vorzugsrichtung der Ionenbewegung mehr. Die Zahl der (A)-Bahnen nimmt ab, wenn das Ionenbündel relativ schräg zur Oberfläche eingeschossen wird. Nach anfänglichem Anwachsen der Zahl der (B)-Bahnen vermindert sich deren Zahl ebenfalls zugunsten der stetig zunehmenden Anzahl der (C)-Bahnen. Die Zahl der Kanalbahnen im Kristall beschränkt sich auf Bahnen zwischen

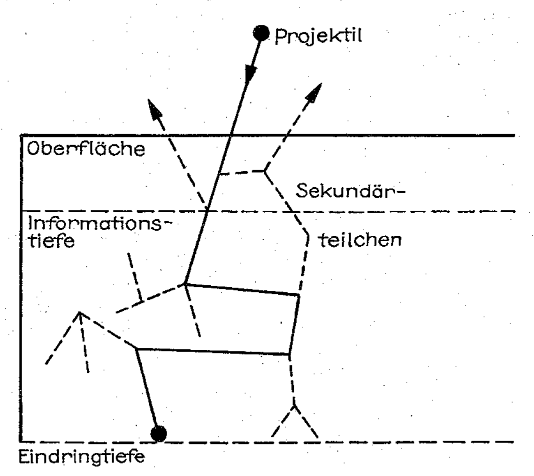


Abb. 6.8.2–13 Ionenbahn im Festkörper

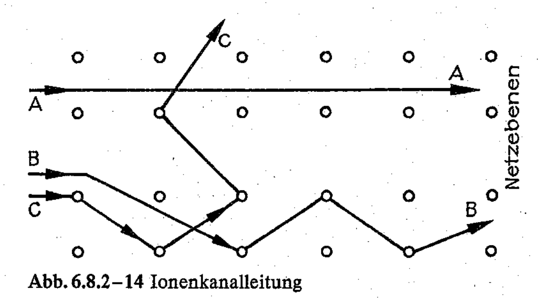


Abb. 6.8.2–14 Ionenkanalleitung

Kristallebenen mit kleinen Millerschen Indizes. Hier ist der Netzebenenabstand am größten, der kleine Atomabstand in der Ebene liefert eine hohe Führungskraft im Kanal.

Ionenzerstäubung (Ionensputtering). Das Bestrahlen des Festkörpers mit Ionen hat zur Folge, daß das Target bis in eine bestimmte Tiefe (Austrittstiefe) hinein als Folge eines Kaskadenprozesses (vgl. Abb. 6.8.2–13) zerstäubt wird. Der *Knock-out-Effekt*, d.h. die Ablösung eines Festkörperatoms durch einen direkten Ionentreffer, ist sehr unwahrscheinlich. Dagegen kommt der *Knock-in-Effekt*, das ist die „Implantation" eines Festkörperatoms in die Tiefe durch einen direkten Ionentreffer, häufiger vor. In der Kollisionskaskade können die Festkörperatome angeregt bzw. (mehrfach) ionisiert werden. Für die Beschreibung der Sekundärteilchenemission durch Ionenbestrahlung gibt es kein allgemein gültiges Modell. Man geht heute davon aus, daß sich längs einer Ionenbahn ein lokaler Plasmazustand bildet. Im (vorausgesetzten) thermodynamischen Gleichgewicht kann über ein Massenwirkungsgesetz das Verhältnis der gebildeten Ionen, Elektronen und Neutralteilchen bestimmt werden. Mit diesem als *Eggert-Saha-Gleichung* bezeichneten Gesetz kann über die Konzentration der emittierten Sekundärionen auf die Konzentration der entsprechenden Elemente im Festkörper geschlossen werden. Das ist die Grundlage der *Sekundärionen-Massenspektroskopie (SIMS)*.

Das Ionenmikroskop. Ein mit einer Ionenoptik ausgestattetes Massenspektrometer gestattet die Abbildung der Festkörperoberfläche im „Licht" der Sekundärionen. Die laterale Auflösung eines solchen Mikroskops liegt bei 1 µm. Der Bildkontrast im Ionenmikroskop entsteht durch Materialkontrast (Konzentrationsunterschiede), Orientierungskontrast, Topografiekontrast (Abschattungseffekt) und durch Energiekontrast (Fokussierungsbedingung der Ionenoptik).

Technische Anwendungen der Ionenimplantation. Das Verfahren der Implantation hat in der Halbleitertechnologie besondere Bedeutung. Durch Dotierung bestimmter Kristallschichten werden die elektronischen Eigenschaften (*pn*-Übergänge) von aktiven Bauelementen (Dioden, Transistoren) erzeugt. Die Implantation kann auch auf Bauelemente in integrierter Form angewendet werden. Bestrahlt man dünne durchsichtige Filme mit leichten Ionen (z. B. Kohlenstoff), so läßt sich eine gezielte ortsabhängige Modulation des Brechungsindex erzeugen. Damit stehen *Lichtleiter* zur Verfügung, die anstelle gedruckter Schaltungen in Rechen- und Steueranlagen den Informationstransport und unter Ausnutzung magnetischer Eigenschaften die Informationsspeicherung übernehmen können. Mit konventionellen Methoden nicht mischbare Elemente lassen sich mit Hilfe der Implantation legieren. Die Oberflächen von Metallen und Legierungen verändern ihre Eigenschaften, wie *Festigkeit, Härte, Korrosionsbeständigkeit* und *katalytische Aktivität*. Eine Ionenbestrahlung von Supraleitern erhöht deren *Sprungtemperatur* sowie den *kritischen Strom* und die *kritische Feldstärke*.

6.8.2.5. Wechselwirkung mit geladenen Kernteilchen. Geladene Kernteilchen klassischer Atome sind Protonen und Deuteronen (vgl. 5.2.), die identisch sind mit den Atomkernen von Wasserstoff (H) und Deuterium (D). Bei einer spontanen Kernumwandlung schwerer Kerne können zwei Protonen und zwei Neutronen emittiert werden, die sich außerhalb des Kerns zu einem α-Teilchen (Heliumion He^{2+}) zusammenfügen. Somit ist die Wechselwirkung des Festkörpers mit geladenen, schweren Kernteilchen (Baryonen) klassischer Atome die Wechselwirkung mit den leichtesten Atomkernen von Wasserstoff, Deuterium und Helium. Die geladenen Kernbestandteile exotischer Atome sind zahlreicher. Neben den klassischen Bausteinen sind das negativ geladene Anti-Protonen, Myonen, Mesonen und Hyperonen, die sich in Masse, Ladung, Spin und in der Lebensdauer voneinander und von den klassischen Teilchen unterscheiden (vgl. 5.1.4.). Die exotischen Teilchen sind kurzlebig. Ihre Lebensdauer reicht von $\approx 10^{-6}$ s für Myonen, bis 10^{-10} s für Hyperonen. Sie lassen sich in ähnlicher Weise wie das Positron (vgl. 6.8.2.3.) und die klassischen Kernbausteine für die Untersuchung von Wechselwirkungseffekten mit der kondensierten Materie einsetzen. Die Betrachtungen zur Ionenwechselwirkung (vgl. 6.8.2.4.) mit dem Festkörper sind hierher übertragbar. Andererseits muß man beachten, daß mit Protonen und α-Teilchen leichte Ionen sehr hoher kinetischer Energie bis zu extrem relativistischen Geschwindigkeiten zur Verfügung stehen. In Gestalt der Myonen gibt es 207mal schwerere „Elektronen". Daraus ergeben sich für die Teilchen de-Broglie-Wellenlängen (vgl. 6.8.2.1.), deren Größe mit dem Radius der Atomkerne vergleichbar werden kann. Schnelle Protonen oder α-Teilchen regen beim Eindringen in den Festkörper dessen Atome an oder ionisieren sie (vgl. 6.8.2.4.). Die benötigte Energie wird der kinetischen Energie der Kernteilchen entzogen. Deren Abnahme durch Wechselwirkung mit den Elektronenhüllen der Festkörperatome ist die *Ionisationsbremsung*. Bei der Relaxation des angeregten Atoms in den Grundzustand wird ein charakteristisches Spektrum elektromagnetischer Strahlung emittiert. Durch die Ionisationsbremsung werden die leichten Kerne wenig gestreut. Sie fliegen nach Passieren einer dünnen Folie nahezu in der Einschußrichtung weiter. Der Energieverlust bei der Ionisationsbremsung wird in erster Näherung durch die Massendicke (ϱr) bestimmt, wobei ϱ die Massendichte und r die Schichtdicke sind. Die Änderung der kinetischen Energie der Kernteilchen mit der Massendicke durchläuft ein Minimum und variiert danach nur noch wenig mit der kinetischen Energie (Abb. 6.8.2.–15).

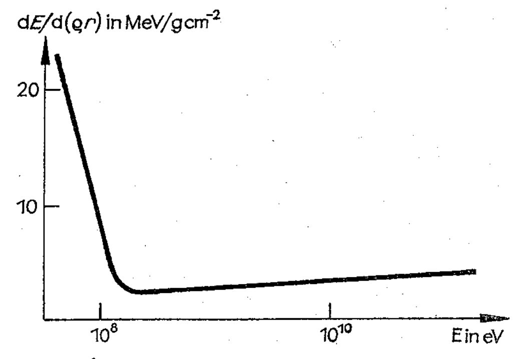


Abb. 6.8.2–15 Ionisationsbremsung

Die sich aus der Ionisationsbremsung ergebende *Eindringtiefe d* (die Reichweite ist $R = \varrho d$, ϱ = Massendichte) hängt ab von der kinetischen Energie der Kernteilchen, von ihrer Masse und vom Einschußwinkel. Bei senkrechtem Auftreffen dringen Protonen der Anfangsenergie 10^7 eV in Festkörpern aus schweren Atomen (Blei) etwa ein 1 mm ein ($L = 10^{-1}$ g/cm²), α-Teilchen der gleichen Energie nur 10^{-2} mm ($L = 10^{-2}$ g/cm²). In diesem Energiebereich dominiert die Ionisationsbremsung. Bei Energien $> 10^8$ eV und

Massendicken $> 10\,g/cm^2$ wird die Eindringtiefe durch andere Energieverluste begrenzt (*Strahlungsbremsung* und *Kerntreffer*).
Eine Teilchenstreuung in den Coulomb-Feldern der Elektronenhülle und des Atomkerns sieht letztere dominieren. Durch den Energieaustausch werden Photonen emittiert. Die Teilchenstreuung im Kernfeld erzeugt ein kontinuierliches Bremsspektrum. Die Strahlungsbremsung ist bei Partikeln, die schwerer sind als Elektronen, gegenüber der Ionisationsbremsung stark benachteiligt. Deshalb ist es möglich, charakteristische Röntgenstrahlen mit einem sehr geringen Streuuntergrund durch Protonen- oder auch Mesonenanregung zu erzeugen. Die Spektren können sich durch Satellitenstrukturen von denen mit Elektronenanregung unterscheiden. Die Strahlungsbremsung ist mit wachsender kinetischer Energie nicht nahezu konstant wie die Ionisationsbremsung (vgl. Abb. 6.8.2–15), sondern wächst etwa proportional mit E weiter an. Die Reichweite kann nicht als Funktion der Kernteilchen- und Festkörpereigenschaften angegeben werden. Nach der Länge L/ϱ (L = *Strahlungslänge*) ist im Mittel die Anfangsenergie der Kernteilchen E_a auf E_a/e abgefallen, es gilt $E = E_a \exp(-\varrho r/L)$. Protonen und α-Teilchen geringer kinetischer Energie werden im Coulomb-Feld des Kerns elastisch gestreut (vgl. 6.8.2.4., Rutherford-Streuung). Neben der Wechselwirkung mit den Feldern von Elektronenhüllen und Kernen kann ein eingeschossenes Kernteilchen einen Atomkern des Festkörpers auch direkt treffen. Infolge ihrer relativ zum Elektron großen Masse zählen die Kerne von Wasserstoff und Helium zu den kernaktiven Partikeln (vgl. 5.2.). Da sie positiv geladen sind, müssen sie das Coulombsche Kernfeld überwinden, bevor sie im Kern des Festkörperatoms eine Kernreaktion auslösen können. Wird der Coulomb-Wall infolge zu geringer kinetischer Energie nicht überwunden, tritt mit der unelastischen Streuung die Strahlungsbremsung (Emission von γ-Strahlung) ein.

6.8.3. Wechselwirkung mit ungeladenen Kernteilchen

Ungeladene Kernteilchen, z. B. Neutronen, werden in den Coulomb-Feldern der Elektronen und der Kerne des unmagnetischen Festkörpers nicht gestreut. Infolge des Spinmomentes des Neutrons kommt es in magnetischen Festkörpern zu elektromagnetischen Wechselwirkungen mit den Elektronen, die für das magnetische Moment verantwortlich sind. Wegen der Ladungsneutralität ist die Kollision mit Festkörperatomen die wichtigste Wechselwirkungsreaktion der Neutronen. Dabei können elastische Reflexion, unelastische Reflexion und ein Eindringen in den Kern auftreten. Bei genügend hoher kinetischer Energie kann das Neutron ein oder mehrere Atome von den Gitterplätzen des Festkörpers entfernen und Leerstellen sowie Zwischengitterplatzbesetzungen (Schottky- und Frenkel-Defekte) erzeugen. Die wichtigsten Informationen, die man aus der Wechselwirkung zwischen Neutronen und kondensierter Materie erhält, sind Spektren der Gitterschwingungen (vgl. 6.4.), der Spinwellen (Magnonen in ferro- und antiferromagnetischen Substanzen) und der Molekülschwingungen (niederenergetische lokalisierte Schwingungen) in Kristallen.
6.8.3.1. Neutronenstreuung. Nach der de-Broglie-Relation haben thermische Neutronen eine den Atomabständen im Kristall entsprechende Wellenlänge ($\approx 0{,}4\,nm$). Die kinetische Energie thermischer (langsamer) Neutronen beträgt $\approx 5 \cdot 10^{-3}\,eV$, ist also $\approx 10^6$mal kleiner als die der Röntgenstrahlung gleicher Wellenlänge. Die geringen Energiebeträge gestatten eine genaue Untersuchung der Atomschwingungen im Festkörper.
Elastische Streuung. Auf den Festkörper wird durch Neutronenstoß keine Energie übertragen. Der Kristall hat gegenüber dem Neutron eine unendlich große Masse. Der auf den Festkörper übertragene Impuls ist $\Delta p = m_N(v_0 - v)$, mit m_N als Neutronenmasse sowie der Geschwindigkeitsdifferenz der Neutronengeschwindigkeit vor und nach der Streuung. Die Beträge der Geschwindigkeiten sind gleich groß. Wegen der periodischen Anordnung der Atome im Kristallverband unterliegt die Impulsübertragung einschränkenden Bedingungen. Der Impuls $\Delta \boldsymbol{p}$ wird auf den gesamten Kristall als Ganzes übertragen. Mit der Einführung von Wellenvektoren $\boldsymbol{k}_0$ und $\boldsymbol{k}$ ist $\Delta \boldsymbol{p} = \hbar(\boldsymbol{k}_0 - \boldsymbol{k})$ nur dann ein vom Kristall aufnehmbarer Impuls, wenn die Differenz der Wellenzahlvektoren einem Vektor des reziproken Kristallgitters (vgl. 6.1.) entspricht. Damit ist die elastische kohärente Neutronenstreuung (*Neutronenbeugung*) analog zur Röntgen-Strahlbeugung (vgl. 6.8.1.8.). Deshalb sind die experimentellen Anordnungen zur Untersuchung der Neutronenbeugung denen der Röntgen-Strukturanalyse (Debye-Scherrer-, Laue- und Drehkristalltechnik) äquivalent.
Jedoch besteht in den Streulängen ein wesentlicher Unterschied zur Röntgen-Beugung. Leichte Atome (kleine Z-Werte) haben nur einen geringen Einfluß auf die Röntgen-Streustrahlung. Im Periodensystem eng benachbarte Elemente unterscheiden sich intensitätsmäßig nur wenig. Der Wirkungsquerschnitt von Atomkernen für die Streuung von Neutronen ändert sich stark mit der Protonen- und Neutronenzahl und hat auch für leichte Kerne große Werte. Deshalb ist die Neutronenstreuung eine wertvolle komplementäre Nachweismethode. Besonders deutlich ist das für Wasserstoff. Durch Röntgen-Strahlstreuung ist in ein Kristallgitter eingebauter Wasserstoff nicht nachweisbar. Dagegen hat Wasserstoff einen sehr großen Wirkungsquerschnitt für Neutronenstrahlung. Die Strukturaufklärung organischer Festkörper gelingt deshalb mit Neutronenstreuung. Der Streuquerschnitt je Raumwinkel- und Energieeinheit (differentieller Streuquerschnitt) addiert sich aus einem kohärenten und einem inkohärenten Anteil. Inkohärente (nicht interferenzfähige) Streuwellen entstehen, wenn verschiedene Isotope bzw. die Orientierung der magnetischen Momente der Atomkerne auf unterschiedlichen Gitterplätzen statistisch verteilt sind. Bei großen Streuquerschnitten kann der isotrope inkohärente Anteil den kohärenten Anteil deutlich übertreffen.
Unelastische Streuung. Die Bestandteile des Wechselwirkungssystems ändern Energie und Impuls. Wenn die Neutronen nicht zu sehr von den Atomkernen eines Festkörpers absorbiert werden, kann deren unelastische Streuung zur experimentellen Bestimmung von Gitterschwingungsspektren verwendet werden. Durch Neutronenwechselwirkung mit den schwingenden Atomkernen des Festkörpers werden Phononen absorbiert oder emittiert (Abb. 6.8.3–1).
Der Streuvorgang umfaßt Einphononen- und Multiphononenprozesse. Die Dispersionsrelation normaler Gitterschwingungen ergibt sich aus der kohärenten Neutronenstreuung durch Einphononenprozesse. Bei diesem Streuvorgang wird die Schwingungsenergie eines Oszillators einer ganz bestimmten Frequenz ω durch ein Neutron um $\hbar\omega$ vergrößert oder vermindert. Dieses Spektrum ist gut strukturiert und relativ leicht auswertbar. Die kohärente Multi-

phononenstreuung und die inkohärente Streuung liefern ein kontinuierliches, gering strukturiertes Streuspektrum. Die Zahl der angeregten Schwingungen unterschiedlicher Frequenzen und Wellenvektoren ist so groß, daß das Wechselwirkungssystem unterbestimmt ist. Die Zahl der für die Auswertung zur Verfügung stehenden Gleichungen kann erhöht werden, indem Neutronenstreuspektren für verschiedene Neutronenenergien und Probenorientierungen registriert werden. Leerstellen und Zwischengitterplatzbesetzungen beeinflussen die Neutronenstreuung. Die Streuverluste nehmen mit steigender Defektkonzentration zu. Schnelle Neutronen erzeugen selbst Defekte im Festkörper. Bei diesen Strahlenschäden lagern sich die Punktdefekte zu Clustern (Mikroporen) zusammen, was zu einem Aufquellen *(Swelling-Effekt)* des bestrahlten Festkörpers führt.

6.8.3.2. Magnetische Wechselwirkungen. Außer dem Wirkungsquerschnitt der Kernwechselwirkung gibt es für Neutronen auch noch einen elektronischen Wirkungsquerschnitt, sofern im streuenden Festkörper ungepaarte Elektronen und folglich magnetische Momente vorhanden sind. Die *elastische kohärente Streuung* in einem magnetischen Kristall gibt über die Anordnung und Verteilung der magnetischen Momente im Festkörper Auskunft. Bei einer *unelastischen Streuung* werden Spinwellen angeregt, deren Quanten die *Magnonen* sind. Das Kristallgitter hat eine periodische Änderung der Spinorientierung. Die umgeklappten Spins repräsentieren einen angeregten Zustand, d.h. sie erhöhen den Energieinhalt des Festkörpers. Der periodischen Änderung der Spinorientierung zugeordnet ist eine periodische Änderung der Elektronendichteverteilung. Diese kann durch unelastische Neutronenstreuung nachgewiesen werden. Die magnetische Neutronenstreuung kann durch äußere Einflüsse, wie Magnetfelder oder Temperaturerhöhung (Curie-Temperatur, Néel-Temperatur), stark modifiziert werden.

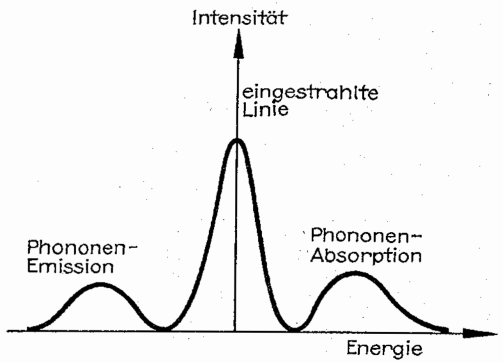


Abb. 6.8.3–1 Spektrum unelastischer Neutronenstreuung

6.9. Hochfrequenzspektroskopie an Festkörpern

Die Untersuchung der Eigenschaften von Festkörpern in einem von außen angelegten HF-Feld (Außenfeld) ist eine der wirkungsvollsten Meßmethoden zur Entschlüsselung der mikroskopischen Eigenschaften kristalliner Substanzen.

Mit Hilfe dieser Methode lassen sich in Metallen und Halbleitern Aussagen zum dynamischen Verhalten der Ladungsträger durch Messung der effektiven Massen m, der Zyklotronmassen m^c und der Streuzeit τ sowie zur Struktur ihres Dispersionsgesetzes $E(\boldsymbol{k})$ gewinnen, indem die Form der Isoenergieflächen durch Messung ihres richtungsabhängigen Durchmessers im $\boldsymbol{k}$-Raum (Fermi-Impuls) oder ihre mittlere Krümmung aus HF-Messungen bestimmt wird. Ähnlich erfolgreich lassen sich HF-spektroskopische Methoden bei der Untersuchung des Verhaltens elementarer

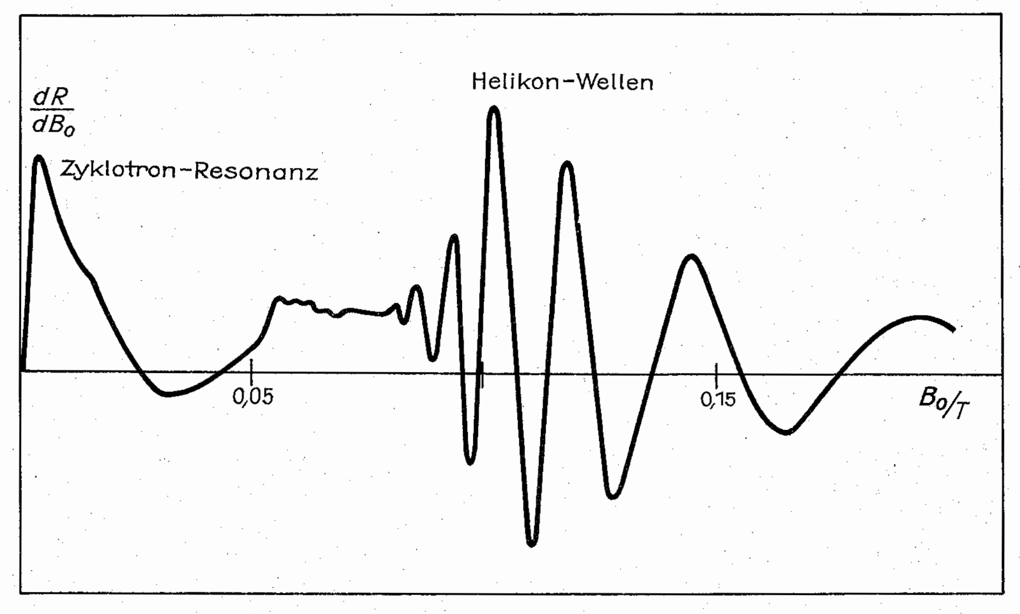


Abb. 6.9.0–1 Meßkurve einer HF-spektroskopischen Untersuchung an einer Parallelplatte aus halbleitendem Bismut-Antimon in einem äußeren Magnetfeld B_0 bei $f_0 = 37$ GHz und $T = 4{,}2$ K

magnetischer Dipole in Festkörpern einsetzen, die Aussagen zur Realstruktur, zu Phasenübergängen, Mikrofeldern, Bindungsverhältnissen und ähnlichem gestatten.
HF-Spektroskopie im eigentlichen Sinn ist die Untersuchung der Frequenzabhängigkeit des HF-Verhaltens einer Probe. Sie ist damit mit der Absorptionsspektroskopie im optischen Bereich vergleichbar. Hauptsächlich wegen der unterschiedlichen Erzeugungs-, Weiterleitungs- und Nachweisverfahren unterteilt man die HF-Spektroskopie nach den *Frequenzbereichen* in

Radiofrequenzspektroskopie $3\cdot10^4\ \text{Hz} < f_0 \leqq 3\cdot10^9\ \text{Hz}$,
Mikrowellenspektroskopie $3\cdot10^9\ \text{Hz} < f_0 \leqq 3\cdot10^{11}\ \text{Hz}$,
Submillimeterspektroskopie $3\cdot10^{11}\ \text{Hz} < f_0 < 10^{13}\ \text{Hz}$.

Die Geometrie des Meßaufbaues wird meist so gewählt, daß auf den *Meßflächen*, d. h. auf den Teilen der Probenoberfläche, auf denen das HF-Feld in der Probe (Innenfeld: E^i, B^i) durch das Außenfeld generiert wird, räumlich konstante HF-Feldstärken vorliegen *(homogenes Anregungsfeld)*.
Das Innenfeld, dessen Verlauf von den Probeneigenschaften und der -geometrie sowie der räumlichen Struktur des Außenfeldes abhängt, regt die Ladungsträger und Kerne über seine elektrische oder magnetische HF-Komponente an, was eine elektrische oder magnetische HF-Polarisierung bzw. HF-Ströme im Material hervorruft. Diese modifizieren in bestimmter Weise das Außenfeld, was mit einer entsprechenden Meßapparatur nachgewiesen wird. In der Mehrzahl der Fälle wird dazu die Dämpfung bzw. die Frequenzverschiebung eines HF-Resonanzkreises durch die Probe ausgenutzt, in dessen HF-Feld sie sich befindet.
Sind die mikroskopischen Anregungsprozesse frequenzabhängig oder werden sie durch andere Parameter beeinflußt, lassen sich Aussagen über die mikroskopischen Eigenschaften des Festkörpers gewinnen. Experimentell ist es schwierig, Untersuchungen mit einer sich über einen größeren Bereich kontinuierlich ändernden Meßfrequenz vorzunehmen, wie es in der optischen Spektroskopie üblich ist. Die meisten Untersuchungen erfolgen deshalb so, daß die HF-Eigenschaften des Festkörpers durch Variation äußerer Parameter, wie Magnetfeld, Temperatur, verändert werden.
Im verallgemeinerten Sinn werden alle parameterabhängigen HF-Untersuchungen, bei denen die Beeinflussung eines HF-Feldes durch eine Festkörperprobe gemessen wird, um Aussagen über mikroskopische Eigenschaften oder Prozesse zu gewinnen, ebenfalls als *HF-Spektroskopie* bezeichnet (Abb. 6.9.0–1).

6.9.1. Mikroprozesse mit Einfluß auf die HF-Eigenschaften

Durch die HF-Feldstärke $\boldsymbol{E}^i$ und die HF-Induktion $\boldsymbol{B}^i$ werden Atomkerne, Ionen, gebundene und Leitungselektronen – sie sind HF-spektroskopisch die Teilsysteme des Festkörpers – zu erzwungenen Schwingungen angeregt. Die Beiträge der einzelnen Teilsysteme zur HF-Polarisierung bzw. zum HF-Strom können von der Anregungsfrequenz f_0, einem statischen Magnetfeld $\boldsymbol{B}_0$, der Temperatur T, aber auch von der Wahl entsprechender Meßbedingungen abhängen, die durch die Polarisations- und Energiestromrichtung des HF-Feldes auf den Meßflächen bzw. die Richtung von $\boldsymbol{B}_0$ zu diesen festgelegt sind. Die HF-Spektroskopie gestattet es daher, die Eigenschaften einzelner Teilsysteme isoliert zu untersuchen.

6.9.1.1. Resonanzeffekte. Als Resonanzen werden alle HF-Effekte bezeichnet, bei denen ein Teilsystem des Festkörpers in einem schmalen Frequenzbereich stark durch das HF-Feld angeregt wird, was eine entsprechende Beeinflussung des Außenfeldes hervorruft.
In klassischer Betrachtungsweise entsteht eine Resonanz, wenn die Anregungsfrequenz f_0 mit einer Eigenfrequenz eines Teilsystems übereinstimmt und die Dämpfung genügend klein ist (vgl. 2.1.5.). Im quantenmechanischen Modell entspricht die Resonanz einem Übergang zwischen zwei Energieniveaus, wenn die Energie des eingestrahlten Photons $\hbar\omega_0$ gleich der Energiedifferenz ΔE der Niveaus ist ($\omega_0 = 2\pi f_0$).
Hängt die Eigenfrequenz von äußeren Parametern ab ($\boldsymbol{B}_0$, T), so kann sie durch entsprechende Parameterwahl auf eine festgehaltene Anregungsfrequenz „geschoben" werden. Starke Änderungen der HF-Eigenschaften der Probe in einem kleinen Parameterbereich (Resonanzgebiet) werden als *Resonanzlinie* bezeichnet.
Im Fall eines starken Magnetfeldes $\boldsymbol{B}_0$ wachsen die Eigenfrequenzen Ω i. allg. mit dem Magnetfeld an. Der Fall *starker Magnetfelder* liegt vor, wenn in dem untersuchten Teilsystem die Bedingung $\Omega\tau \gg 1$ erfüllt ist (τ = Streuzeit).
Magnetische Resonanzen. Durch die Larmor-Präzession magnetischer Dipole μ_I (vgl. 2.1.7.3.) um $\boldsymbol{B}_0$ besitzt das Dipolmoment einen oszillierenden Anteil senkrecht zu $\boldsymbol{B}_0$ (Abb. 6.9.1–1).
Wird ein HF-Feld angelegt, dessen magnetische Komponente $\boldsymbol{B}^i$ senkrecht zu $\boldsymbol{B}_0$ (zirkulär) polarisiert ist, so erhöht sich im Resonanzfall ($\Omega = \omega_0$, $\Omega\tau \gg 1$) über die Kräfte, die $\boldsymbol{B}^i$ auf μ_I ausübt, die Präzessionsenergie.
Zyklotronresonanzen sind die zweite wesentliche Gruppe von Resonanzeffekten in einem statischen Magnetfeld. Sie hängen mit der Ablenkung freibeweglicher Ladungsträger im Festkörper (Elektronen, Löcher) durch die Lorentz-Kraft $\boldsymbol{F} = e[\boldsymbol{v} \times \boldsymbol{B}_0]$ zusammen (vgl. 2.1.3., $\boldsymbol{v}$ = Teilchenge-

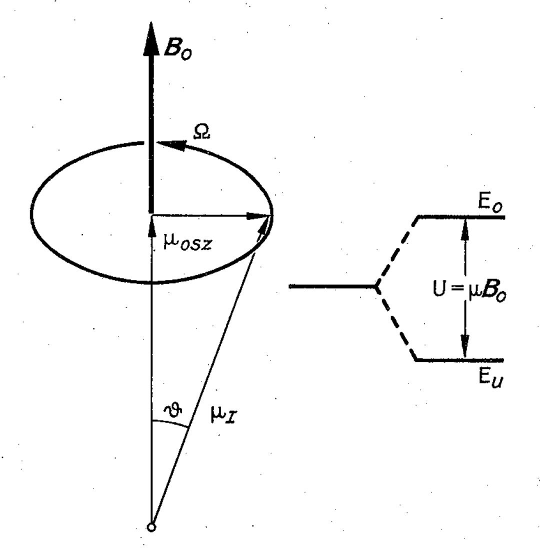


Abb. 6.9.1–1 Larmor-Präzession eines magnetischen Dipols mit dem zugehörigen Energieniveauschema für $I = 1/2$ (Ω = Kreisfrequenz, E_0, E_u = oberes bzw. unteres Energieniveau, μ_I, μ_{osz} = Gesamt- bzw. oszillierender Dipolbetrag, ϑ = Präzessionswinkel)

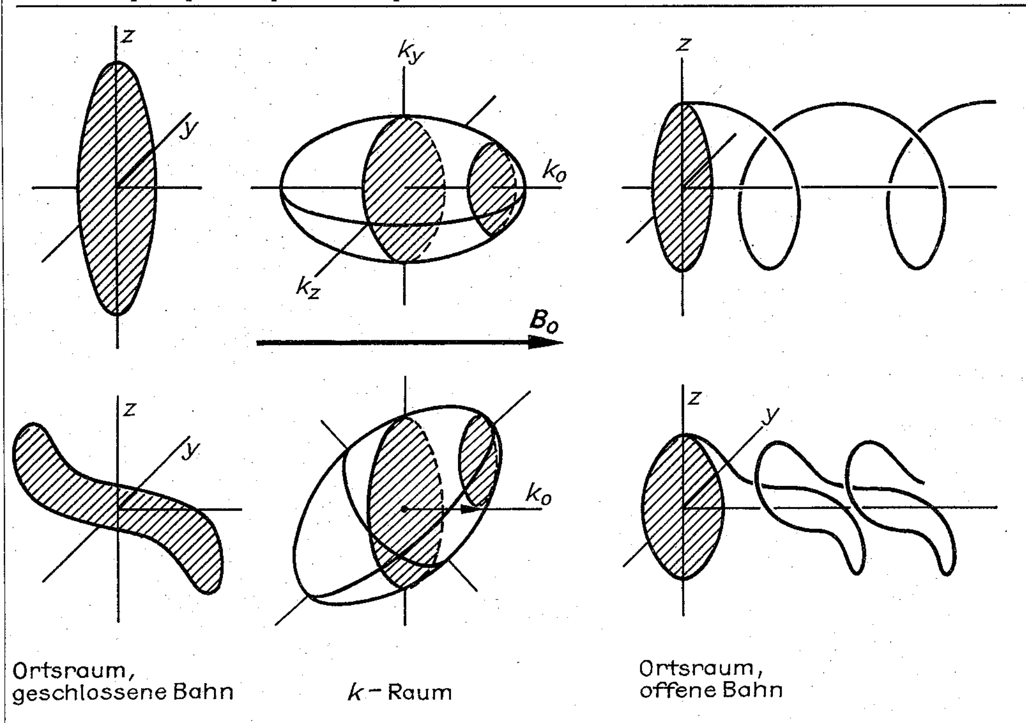


Abb. 6.9.1-2 Umlaufbahnen im Orts- und k-Raum

schwindigkeit, e = Elementarladung). Im Fall starker Magnetfelder durchlaufen die Ladungsträger periodische Umlaufbahnen. Ihr Impuls p_0 in Magnetfeldrichtung ist eine Erhaltungsgröße (Abb. 6.9.1-2). Die *Zyklotronfrequenz* (Umlauffrequenz um das Magnetfeld) $\Omega = eB_0/m^c$ ist proportional dem Magnetfeld. Im Resonanzfall $\Omega = \omega_0$, $\Omega\tau \gg 1$ werden die Ladungsträger durch ein geeignet angelegtes HF-Feld über die HF-Feldstärke E^i auf der Umlaufbahn periodisch beschleunigt.

6.9.1.2. Wechselwirkung mit dem Ladungsträgerplasma. Sind im Festkörper freibewegliche Ladungsträger, führt die elektrische HF-Komponente durch entsprechende Beschleunigungen zu Strömen und Raumladungen, also zu kollektiven Bewegungen ähnlich wie in einem Gasplasma.

Die *HF-Leitfähigkeit* σ, die das kollektive Verhalten der Ladungsträger im elektrischen Feld beschreibt – sie stellt den Zusammenhang zwischen der Feldstärke E^i und der Stromdichte $j(r)$ her – wird zum bestimmenden Materialparameter. Die HF-Leitfähigkeit wird dann komplex. Der Imaginärteil beschreibt das nichtdissipative Verhalten.

Der *Fall hoher Frequenzen* liegt vor, wenn $\omega_0\tau \gg 1$ ist, d. h. die Streuzeit der Teilchen τ viel größer als die Periodendauer der Anregungsfrequenz ist. Der Ladungsträger vollführt dann zwischen zwei Stößen mehrere Schwingungen, die nicht mit einer Energiedissipation verbunden sind und zu einer HF-Polarisierung $\boldsymbol{P}$ führen.

In starken Magnetfeldern, in denen durch die Ablenkung und Lokalisierung auf Schraubenbahnen die HF-Leitfähigkeit modifiziert wird, ergeben sich Besonderheiten, die die Ursache für kollektive Schwingungen und schwach gedämpfte elektromagnetische Wellen im Plasma sind.

Untergrund. Das HF-Feld polarisiert auch die Teilsysteme, die kein resonantes bzw. Plasmaverhalten zeigen (z. B. Ionenrümpfe). Diese Wechselwirkung, die keine signifikanten HF-Besonderheiten verursacht, aber das Innenfeld modifiziert, faßt man zur *Gitterdielektrizitätskonstanten* ε_L und *Gitterpermeabilität* μ_L zusammen.

6.9.2. Das HF-Feld im Innern der Probe

In der HF-Spektroskopie hat man es immer mit Proben endlicher Abmessung zu tun, die an ein äußeres HF-Feld angekoppelt werden und es entsprechend modifizieren. Um zu quantitativen Aussagen zu gelangen, ist es notwendig, die *Meßbedingungen* zu kennen, d. h. die Struktur des HF-Feldes in der Probe als Funktion der Materialparameter, die ihrerseits von äußeren Parametern und den mikroskopischen Prozessen bestimmt werden. Ebenso wichtig ist der Zusammenhang zwischen Außen- und Innenfeld.

6.9.2.1. Die Maxwell-Gleichungen in der Probe. Das HF-Feld in der Probe gehorcht den Maxwell-Gleichungen, durch deren Lösung sich die Feldstärkeverteilung berechnen läßt. Mit einer sinusförmigen Zeitabhängigkeit der in die Maxwellgleichungen eingehenden Felder, z. B. $\boldsymbol{E}^i = \boldsymbol{E}^i(\boldsymbol{r})\exp(i\omega_0 t)$, was bei der als monochromatisch vorausgesetzten Anregung bei kleinen Feldstärken gilt, und durch Vorgabe einer ebenen Meßfläche mit einem homogenen Anregungsfeld, wodurch die Feldgrößen nur noch von der Ortskoordinate senkrecht zur Meßfläche abhängen, kann eine bedeutende Vereinfachung der Maxwell-Gleichungen erreicht werden. Die x-Koordinate liegt senkrecht zur Meßfläche. Gl. 6.9.2-1:

$dE^i_z/dx = i\omega_0 B^i_y$, $dE^i_y/dx = -i\omega_0 B^i_z$, $B^i_x = 0$, $dB^i_x/dx = 0$,
$dH^i_y/dx = j_z + i\omega_0 D^i_z$, $-dH^i_z/dx = j_y + i\omega_0 D^i_y$,
$j_x + i\omega_0 D^i_x = 0$, $dD^i_x/dx = 0$.

6.9.2.2. Materialparameter und HF-Feld. Das HF-Feld wird entscheidend über den Zusammenhang zwischen den einzelnen Feldgrößen ($\boldsymbol{E}^i$ = elektrische Feldstärke, $\boldsymbol{H}^i$ = magnetische Feldstärke, $\boldsymbol{D}^i$ = dielektronische Verschiebung, $\boldsymbol{B}^i$ = magnetische Induktion des HF-Feldes in der Probe), der als linear vorausgesetzt werden kann, beeinflußt. Im HF-Bereich sind die HF-Leitfähigkeit σ, die Dielektrizitätsfunktion ε und die Permeabilität μ i. allg. Tensoren mit komplexen Elementen. Den Zusammenhang mit den untersuchten mikroskopischen Prozessen liefert die Theorie des jeweiligen Effektes.
Wird vorausgesetzt, daß die Materialparameter nicht von der räumlichen Struktur des HF-Feldes abhängen, d.h. wellenlängenunabhängig sind, läßt sich Gl. 6.9.2–1 in die Form Gl. 6.9.2–2

$$-d^2 E^i_\alpha(x)/dx^2 = \mu_L (i\omega_0 \sigma_{\alpha\beta} + \omega_0^2 \varepsilon_L) E^i_\beta(x), \quad \alpha = y, z$$
$$0 = (\sigma_{x\beta} + i\omega_0 \varepsilon_L) E^i_\beta(x), \quad \beta = x, y, z$$

bringen. Über zweifach vorkommende Indizes ist zu summieren. Die Lösung von Gl. 6.9.2–2 sind in x-Richtung laufende, exponentiell gedämpfte ebene Wellen. Gl. 6.9.2–3

$\boldsymbol{E}^i(x, t) \sim \exp(i(\omega_0 t - q^* x))$ mit $q^* = q' - i/\delta$.

q^* ist der komplexe Wellenvektor in x-Richtung, der mit der Wellenlänge des HF-Feldes im unendlichen Material über $q' = 2\pi/\lambda$ und mit der Dämpfungs- oder Abschirmlänge über seinen Imaginärteil verbunden ist. Setzt man Gl. 6.9.2–3 in Gl. 6.9.2–2 ein, so besitzt das homogene Gleichungssystem nur für bestimmte q^* eine Lösung (aus der Säkulargleichung). q^* wird eine Funktion der Materialparameter und der Frequenz.
In zwei Fällen läßt sich das Innenfeld allein aus der Größe des Wellenvektors q^* bestimmen:
Räumlich konstantes HF-Feld. Sind die Probenabmessungen viel kleiner als die Wellen- und Dämpfungslänge (im unendlichen Material), kann die räumliche Änderung des HF-Feldes in der Probe vernachlässigt werden. Da im Normalfall im Material die Wellenlänge kürzer als im Vakuum ist, ist das Außenfeld um die Probe räumlich konstant, wodurch sich der Zusammenhang zwischen Innen- und Außenfeld ähnlich wie in der Elektrostatik ergibt.
Wellenfeld. Ist die Wellenlänge λ des HF-Feldes in der Probe viel kleiner als die Dicke d in Ausbreitungsrichtung und die Welle nur schwach gedämpft ($\delta \gg d$), breiten sich im Material elektromagnetische Wellen aus, deren räumliche Struktur durch den Wellenvektor bestimmt ist.
6.9.2.3. Spezielle Lösungen für das Festkörperplasma. Während in Isolatoren fast immer der Fall des räumlich konstanten bzw. des Wellenfeldes realisiert werden kann, bestimmt in Leitern der Zusammenhang zwischen HF-Strom und -Feldstärke entscheidend den Feldverlauf.
Normaler Skineffekt. Unter *lokalen Bedingungen* klingt durch die Wirkung der in der Oberfläche induzierten HF-Ströme das HF-Feld in der Probe exponentiell ab. Lokal bedeutet, daß die freie Weglänge der Plasmateilchen l viel kleiner als die charakteristischen Längen des HF-Feldes (λ, δ) ist und die Ladungsträger damit zwischen zwei Stößen in einem räumlich konstanten Feld beschleunigt werden. Ihr Beitrag zur (komplexen) Leitfähigkeit hängt nicht von der Struktur des HF-Feldes ab.

Im einfachsten Fall eines einkomponentigen Plasmas, dessen Ladungsträger eine isotrope Masse besitzen, gilt $\sigma(\omega_0) = \sigma_0/(1 + i\omega_0\tau)$ mit der *Gleichstromleitfähigkeit* σ_0, $\sigma_0 = ne^2\tau/m^*$ (τ = Streuzeit, n = Ladungsträgerkonzentration, m^* = effektive Masse).
Unter Berücksichtigung der Oberfläche (Meßfläche bei $x = 0$) erhält man für die Feldstärke in der Probe

$\boldsymbol{E}^i(x) = [\boldsymbol{E}'(0)/iq^*] \exp(-iq^* x)$ (Gl. 6.9.2–4).

$\boldsymbol{E}'(0) = \lim_{x\to 0} (d\boldsymbol{E}/dx)$ wird durch das HF-Magnetfeld auf der Meßfläche bestimmt (vgl. Gl. 6.9.2–1).
Für niedrige Frequenzen [$\omega_0\tau \ll 1$, $\sigma(\omega_0)$ wird σ_0] erhält der Wellenvektor dann die Form $q^* = 2\pi/\lambda - i/\delta = \sqrt{\omega_0^2 \mu_L \varepsilon_L (1 - i\sigma_0/\varepsilon_L\omega_0)} \approx \sqrt{\omega_0 \mu_L \sigma_0/2}\,(1 - i)$, was in Leitern eine gute Näherung ist, da dort fast immer $\sigma_0 \gg \varepsilon_L \omega_0$ gesetzt werden kann.
$\delta = (\omega_0 \mu_L \sigma_0/2)^{-1/2}$ ist die *normale Skintiefe*. Da $\delta < \lambda$ ist, bestimmt in Gl. 6.9.2–4 das exponentielle Abklingen in einer dünnen Oberflächenschicht, der *Skinschicht*, den Verlauf des HF-Feldes.
Für hohe Frequenzen [$\omega_0\tau \gg 1$, $\sigma(\omega_0)$ wird $\sigma_0/i\omega_0\tau$] erhält der Wellenvektor die Form Gl. 6.9.2–5

$$q^* = \sqrt{\omega_0^2 \mu_L \varepsilon_L - \mu_L \sigma_0/\tau} = \sqrt{\mu_L \varepsilon_L (\omega_0^2 - \omega_p^2)}\,.$$

Dieser Fall hat im HF-Bereich nur für entsprechend dotierte Halbleiter Bedeutung, da sich in Metallen für $\omega_0\tau \gg 1$ wegen der kleinen Skintiefe und der hohen Ladungsträgergeschwindigkeit keine lokalen Bedingungen herstellen lassen. Für Frequenzen unterhalb der *Plasmafrequenz* $\omega_p = \sqrt{(ne^2/m^* \varepsilon_L)}$ ist q^* rein imaginär, d. h. das HF-Feld in der Probe klingt exponentiell ab. Ist die Meßfrequenz größer als ω_p, breiten sich elektromagnetische Wellen im Plasma aus.
Anomaler Skineffekt. Ein völlig veränderter Abschirmmechanismus liegt in Metallen (hohe Reinheit, tiefe Temperaturen) vor, wenn die freie Weglänge der Ladungsträger viel größer als die Skintiefe wird. Den Abschirmstrom erzeugen hier nur die Ladungsträger, die sich in der Skinschicht etwa parallel zur Oberfläche bewegen. Die übrigen sind wegen der großen Geschwindigkeit (Fermi-Geschwindigkeit) nur sehr kurz in der Skinschicht und tragen kaum zum HF-Strom bei. Für den Beitrag des einzelnen Ladungsträgers zur Leitfähigkeit ist entscheidend, wie der Feldstärkeverlauf auf seiner freien Weglänge ist *(nichtlokale Bedingungen)*. Mathematisch läßt sich das durch eine Wellenlängenabhängigkeit des HF-Leitfähigkeitstensors $\hat{\sigma} = \hat{\sigma}(\omega_0, q)$ beschreiben, was die Lösung von Gl. 6.9.2–1 bedeutend verkompliziert.
Sie muß in Form einer Überlagerung (Integration) von ebenen Wellen mit unterschiedlicher Wellenlänge $\lambda = 2\pi/q$ und der zu ermittelnden Amplitude $e(q)$ bestimmt werden, ähnlich wie sich zeitabhängige Funktionen aus den Amplituden der Spektralkomponenten in der Fourierreihe erzeugen lassen. Gl. 6.9.2–6:

$$\boldsymbol{E}^i(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e(q) \exp(-iqx)\, dq$$
$$= -\frac{\boldsymbol{E}'(0)}{\pi} \int_{-\infty}^{\infty} \{\exp(-iqx)$$
$$[q^2 - \omega_0 \mu_L \varepsilon_L (1 - i\,\sigma(\omega_0, q)/\varepsilon_L \omega_0)]\}\, dq\,.$$

Das ist ein universeller Ausdruck für isotrope Leiter, der den Verlauf von $\boldsymbol{E}^i$ in der Probe liefert.

Im einfachsten Grenzfall für Metalle ($\omega_0\tau \gg 1$; λ, $\delta \ll l$; isotrop) gilt etwa $\sigma(\omega_0, q) \approx \mathrm{i}\, a\, \sigma_0/(lq)$, ($a$ = Materialparameter, $a \approx 1$).
Der Feldstärkeverlauf in der Probe ergibt sich damit zu Gl. 6.9.2–7:

$$E^{\mathrm{i}}(x) = E'(0)\,(1/\sqrt{3} - \mathrm{i}/3)\,\delta_{\mathrm{a}} \exp\{-\mathrm{i}(1 - \mathrm{i}\sqrt{3})\,x/\delta_{\mathrm{a}}\}\,.$$

$\delta_{\mathrm{a}} = (a\sigma_0\,\omega_0\,\mu_{\mathrm{L}}/l)^{-1/3}$ ist die *anomale Skintiefe.*

6.9.2.4. Oberflächenimpedanz. Zur Berechnung des Zusammenhanges zwischen dem Innen- und Außenfeld einer leitenden Probe mit Skineffekt kann ausgenutzt werden, daß die Tangentialkomponenten der HF-Feldstärken $E_{\mathrm{t}}^{\mathrm{i}}$ und $H_{\mathrm{t}}^{\mathrm{i}}$ an der Meßfläche ($x = 0$) stetig übergehen. Gleichzeitig legen sie über den Poynting-Vektor S die in die Probe einströmende Leistung fest, $W \sim \mathrm{Re}\,(S \cdot n) \sim \mathrm{Re}\,(E_{\mathrm{t}}(0)\,[H_{\mathrm{t}}(0)\,\mathrm{x}\,n])$ (n = Einheitsvektor in x-Richtung, Re = der Realteil des Spatproduktes).
Aus dem Feldstärkeverlauf für den lokalen (vgl. Gl. 6.9.2–4) und den nichtlokalen Fall (vgl. Gl. 6.9.2–7) unter Berücksichtigung, daß $\mathrm{d}E_{\mathrm{t}}^{\mathrm{i}}/\mathrm{d}x = -\mathrm{i}\,\omega_0\,\mu_{\mathrm{L}}[H_{\mathrm{t}}^{\mathrm{i}}\mathrm{x}\,n]$ ist (vgl. Gl. 6.9.2–1), ist zu entnehmen, daß $E_{\mathrm{t}}^{\mathrm{i}}$ und $H_{\mathrm{t}}^{\mathrm{i}}$ proportional zueinander sind. Die Proportionalitätskonstante – sie hängt nur von den Materialparametern ab – wird als *Oberflächenimpedanz* bezeichnet, $E_{\mathrm{t}}^{\mathrm{i}}(0) = Z^*[H_{\mathrm{t}}^{\mathrm{i}}(0)\,\mathrm{x}\,n]$, wenn ihr Wert an der Meßfläche bestimmt wird. Im allg. Fall ($\hat{\sigma}$ ist dann anisotrop) ist Z^* ein komplexer Tensor der Form $\hat{Z}^* = \hat{R} + \mathrm{i}\hat{X}$. Unter Skineffekt-Bedingungen wird $\hat{Z}^*$ die Meßgröße für die HF-Eigenschaften der Probe. Bei leitenden Proben wird durch die Meßapparatur auf der Meßfläche $H_{\mathrm{t}}(0)$ vorgegeben. Ist der Festkörper isotrop, wird die Energieabsorption $W \sim \mathrm{Re}\,\{Z^*[H_{\mathrm{t}}(0)]^2\}$.
Unter lokalen Bedingungen ist $Z^* = \mu_{\mathrm{L}}\,\omega_0/q^*$ (vgl. Gl. 6.9.2–4) und unter nichtlokalen $Z^* = (1/3)\,\mu_{\mathrm{L}}\,\omega_0\delta_{\mathrm{a}} \times (1 + \sqrt{3}\,\mathrm{i})$ (vgl. Gl. 6.9.2–7). Der Vorteil der Oberflächenimpedanz besteht darin, daß sie Aussagen über die Beeinflussung des Außenfeldes durch die Probe gestattet, ohne das Innenfeld zu kennen.

6.9.3. Elektromagnetische Wellen im Festkörperplasma

Unter Einwirkung eines starken Magnetfeldes B_0 kann die Abschirmwirkung des Plasmas aufgehoben werden, so daß HF-Energie durch eine massive leitende Probe transportiert wird. Als *Einsatzkante* bezeichnet man das Magnetfeld, ab dem bei vorliegenden Anregungsbedingungen eine Wellenausbreitung möglich wird. Die Wellenausbreitung erfolgt senkrecht zur Meßfläche.
Meßgröße bei HF-spektroskopischen Untersuchungen sind die Wellenlänge und die Dämpfung der Wellen in Abhängigkeit vom Magnetfeld B_0 und der Anregungsfrequenz f_0. Über das *Dispersionsgesetz*, d. h. den Zusammenhang $q^* = q^*(\omega_0, B_0)$, sind sie mit den mikroskopischen Parametern, wie effektive Massen, Streuzeiten usw., verknüpft.

6.9.3.1. Magnetoplasmawellen unter lokalen Bedingungen. Lokal bedeutet, daß der Ladungsträger auf seiner Umlaufbahn ein räumlich konstantes HF-Feld „sieht". Ladungsträger, die sich parallel zu B_0 bewegen, werden nicht abgelenkt. Die anderen durchlaufen Schraubenbahnen mit dem Durchmesser D (vgl. 6.9.1.1.).
Zeigt das Magnetfeld B_0 senkrecht zur Meßfläche und damit parallel zum Wellenvektor q^* (Ausbreitungsrichtung) – diese Meßkonfiguration wird als *Faraday-Konfiguration* (Abb. 6.9.3.–1) bezeichnet – herrschen so lange lokale Bedingungen, bis die freie Weglänge der unabgelenkten Ladungsträger, sie bewegen sich am weitesten in Ausbreitungsrichtung der Welle, vergleichbar mit der Wellenlänge wird (lokal: $l \ll \lambda$).
Zeigt das Magnetfeld B_0 parallel zur Meßfläche und damit senkrecht zum Wellenvektor – diese Meßanordnung wird als *Voigt-Konfiguration* (Abb. 6.9.3–2) bezeichnet – wird der Maximaldurchmesser der Schraubenbahnen D zur bestimmenden Länge, der mit der Wellenlänge λ zu vergleichen ist (lokal: $D \ll \lambda$).
Typische Magnetoplasmawellen, die nur unter lokalen Bedingungen ausbreitungsfähig sind (nichtlokale Verhältnisse führen zu Dämpfungen), sind die Helikonen und Alfvén-Wellen.

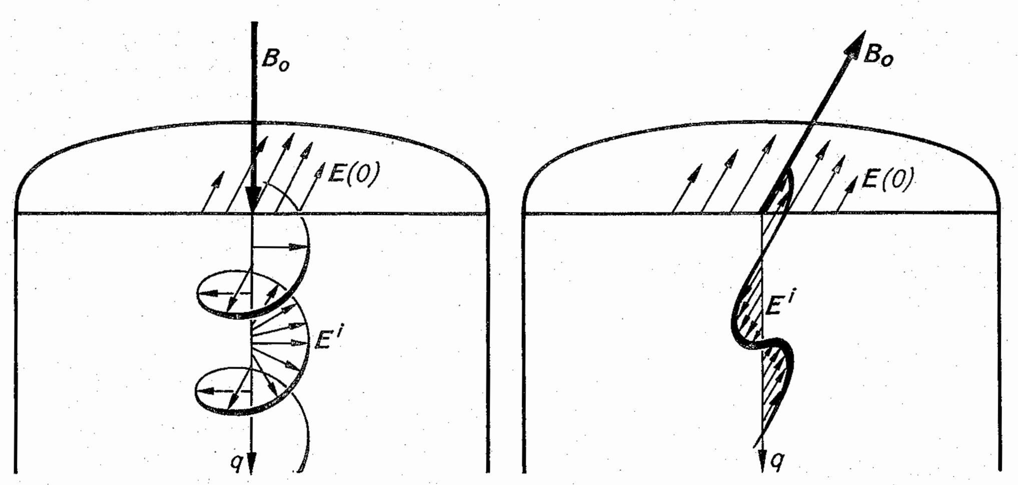


Abb. 6.9.3–1 (links) Faraday-Konfiguration; dargestellt ist die Anregung von Helikonen ($E(0)$ = E-Vektoren des Anregungsfeldes auf der Oberfläche, E^{i} = E-Vektor für die jeweilige Schicht
Abb. 6.9.3–2 (rechts) Voigt-Konfiguration; dargestellt ist die Anregung der ordentlichen Mode ($E(0)$ = E-Vektoren des Anregungsfeldes auf der Oberfläche, E^{i} = E-Vektor für die jeweilige Schicht)

Helikonen sind zirkular oder elliptisch polarisierte elektromagnetische Transversalwellen, die sich vorzugsweise in der Faraday-Konfiguration in Festkörperplasmen ausbreiten, wenn $\boldsymbol{B}_0$ so groß ist, daß die Zyklotronfrequenz Ω der Ladungsträger größer als die Anregungsfrequenz ω_0 ist (Einsatzkante) und die Bedingung $\Omega\tau \gg 1$ erfüllt wird (geringe Dämpfung). Die Plasmen bestehen entweder aus Elektronen oder Löchern (Alkalimetalle bzw. Fremdhalbleiter). Helikonen können bei sehr geringen Meßfrequenzen angeregt werden.

Das *Dispersionsgesetz für Helikonen* läßt sich aus Gl. 6.9.2–1 bei Vorgabe des Leitfähigkeitstensors sowie ε_L und μ_L berechnen. Für ein einkomponentiges Plasma mit isotroper effektiver Masse unter lokalen Bedingungen und B_0 in x-Richtung besitzt er die Form Gl. 6.9.3–3

$$\sigma_{yy} = \sigma_{zz} = \frac{ne^2}{\mathrm{i}m\tilde{\omega}} \frac{1}{1-(\Omega/\tilde{\omega})^2}, \quad \sigma_{xx} = \frac{ne^2}{\mathrm{i}m\tilde{\omega}}$$

$$\sigma_{yz} = -\sigma_{zy} = \frac{ne^2}{m\tilde{\omega}} \frac{(\Omega/\tilde{\omega})}{1-(\Omega/\tilde{\omega})^2}, \quad \tilde{\omega} = \omega_0 - \mathrm{i}/\tau.$$

Die Helikonen verdanken ihre Existenz der Tatsache, daß in der zum Magnetfeld senkrechten Ebene durch die Teilchenablenkung im Magnetfeld $\boldsymbol{B}_0$, ähnlich wie beim Hall-Effekt, HF-Strom und HF-Feldstärke nicht parallel und außerdem zueinander phasenverschoben sind. Die Stromkomponente quer zur Feldstärke wird durch σ_{yz} bestimmt.

Mit dem Ansatz der Gl. 6.9.2–3 wird Gl. 6.9.2–2 in ein lineares Gleichungssystem mit von den Materialparametern ε_L, μ_L und σ sowie $\boldsymbol{q}^*$ abhängenden Koeffizienten überführt, aus dem $\boldsymbol{q}^*$ bestimmt wird:

$$q^{*2} = \omega_0^2 \varepsilon_L \mu_L - \mu_L \omega_0 (\mathrm{i}\sigma_{yy} \mp \sigma_{yz}) \approx \mu_L \varepsilon_L \omega_0^2 \mp \mu_L \omega_0 ne \times [1 \pm \mathrm{i}/(\Omega\tau)]/\boldsymbol{B}_0.$$

Das ist unter Ausbreitungsbedingungen ($\Omega\tau \gg 1$, $\Omega > \omega_0$), eine gute Näherung.

Für Helikonen ist typisch, daß die beiden Feldstärkekomponenten E_y^i und E_z^i zeitlich um $\pi/2$ phasenverschoben sind, was die Rotation des HF-Feldstärkevektors hervorruft (vgl. Abb. 6.9.3–1), und das rotierende Magnetfeld, verglichen mit den Verhältnissen im Vakuum ($E_{\text{Vak}}/B_{\text{Vak}} = c_0$) stark vergrößert ist ($E^i/B^i = \omega_0/q' \ll c_0$).

Alfvén-Wellen sind linear polarisierte Wellen, die sich in Materialien mit gleicher Elektronen- und Löcherkonzentration ($n = p$, *kompensiertes Plasma*) in starken Magnetfeldern ($\Omega\tau > 1$) unter beliebigen Winkeln zu diesem ausbreiten. Um Wellenausbreitung zu ermöglichen und damit die Blindanteile in σ überwiegen, muß die Frequenz ω_0 genügend groß sein ($\omega_0\tau \gg 1$), da die σ_{yz}-Komponente wegen $n = p$ klein ist. Die Einsatzkante liegt zwischen der Zyklotronresonanz der Löcher und Elektronen (vgl. 6.9.4–2). Derartige Wellen können bei tiefen Temperaturen in reinen Halbmetallen (Bi, Sb) und Graphit bei Mikrowellenfrequenzen ($\omega_0\tau > 1$) beobachtet werden.

Das *Dispersionsgesetz der Alfvén-Wellen* hat unter Vernachlässigung des Verschiebungsstromes die Form

$$q'^2 = \left[\omega_0^2 m_0/c_0^2 \varepsilon_0\right] f(n, m^+) B_0^{-2},$$

$f(n, m^+)$ ist die Massendichtefunktion, die von der Ladungsträgerkonzentration und den reduzierten effektiven Massen $m^+ = m^*/m_0$ (m_0 = Masse des freien Elektrons, m^* = effektive Masse) abhängt. Im einfachsten (in der Realität nicht vorkommenden) Fall isotroper Elektronen-(m_n^+) und Löchermassen (m_p^+) ist $f(n, m^+) = n(m_n^+ + m_p^+)$.

6.9.3.2. Zyklotronwellen. Liegen in einem Metall für $B_0 \neq 0$ nichtlokale Bedingungen vor, beruht der Transport von HF-Energie aus der Skinschicht darauf, daß für $\boldsymbol{B}_0$ parallel zur Oberfläche die Teilchen über gekrümmte Bahnen (Krümmungsradius R_c) ihre Anregungsenergie in das Innere tragen (vgl. Abb. 6.9.4–1). Für die Wellenausbreitung sind die Teilchen wesentlich, die in der Skinschicht periodisch beschleunigt werden, was nur in der Nähe der Azbel-Kaner-Zyklotronresonanz der Fall ist (vgl. 6.9.4.2.) und zu HF-Strömen im abgeschirmten Inneren führt, die die Ursache für die Wellenausbreitung sind. Das Dispersionsgesetz ist sehr kompliziert. Ausbreitungsfähig sind derartige Wellen nur in einem geringen Magnetfeldbereich oberhalb der Azbel-Kaner-Resonanzen, in deren Nähe zwei Wellenmoden, der kurzwellige (Wellenlänge kleiner als der Durchmesser des Zyklotronumlaufs) und der langwellige Zweig, ausbreitungsfähig sind. Zyklotronwellen wurden an Alkalimetallen und Wismut beobachtet.

6.9.3.3. Spinwellen. Spinwellen können sich in Festkörperplasmen und Isolatoren ausbreiten, wenn zwischen den Spinmomenten eine *Austauschwechselwirkung* besteht. In ferromagnetischen Substanzen (Leiter und Isolatoren) sind diese Anregungen in Magnetfeldern, die zur Sättigung führen, sehr kurzwellig, so daß dort auch nichtlokale Bedingungen vorliegen. In nichtmagnetischen Materialien mit Spinkopplung der Leitungselektronen sind derartige Wellen mit der paramagnetischen Elektronenresonanz verknüpft (vgl. 6.9.4.1.) und breiten sich nur in unmittelbarer Nähe der Resonanzmagnetfeldstärke aus.

Im klassischen Bild erfolgt der Transport der HF-Energie über den Austausch der Präzessionsenergie zwischen den Spins. Die Welle entsteht durch eine räumlich periodische Drehung des Präzessionswinkels ϑ zwischen 0° und 180° (vgl. Abb. 6.9.1–1).

6.9.3.4. Aus Magnetoplasmawellen ableitbare Materialparameter. Aus Helikonen lassen sich Aussagen über Ladungsträgerkonzentrationen, die Gitterdielektrizitätskonstante und über Streuzeiten gewinnen. Ein weiteres Anwendungsgebiet besteht darin, über die durch das Magnetfeld einstellbare Phasengeschwindigkeit $c = \omega_0/q'$ in der Nähe der Schallgeschwindigkeit Helikonen mit Phononen koppeln zu lassen, und dadurch Aussagen über das Phononenspektrum zu erhalten. Ferner lassen sich aus Besonderheiten des Helikondispersionsgesetzes in der Nähe der Zyklotronresonanz (Einsatzkante) die Zyklotronmassen abschätzen.

Für Alfvénwellen ist die wesentlichste Meßgröße die Massendichtefunktion, die nur für einfachste Energiebänder berechnet werden kann.

Zyklotron- und Spinwellen dienen vor allem zur Charakterisierung der Plasmawechselwirkung. Aus Zyklotronwellen können Aussagen zur Anisotropie einfacher Fermiflächen abgeleitet werden.

6.9.4. Resonanzeffekte

Die Untersuchung der Resonanzeffekte (vgl. 6.9.1.1.) in einem Festkörper besonders unter Einwirkung eines Magnetfeldes ist das wichtigste Teilgebiet der HF-Spektroskopie, da sich aus ihnen die genauesten Aussagen über viele mikroskopische Zusammenhänge ableiten lassen.

Zeitliche Resonanzen sind alle die Effekte, bei denen im klassischen Bild eine phasenrichtige Beschleunigung durch das HF-Feld auf einer periodischen Umlaufbahn erfolgt. Dazu gehören alle Formen der magnetischen Resonanzen,

der Zyklotronresonanz und die magnetischen Oberflächenzustände.

Räumliche Resonanzen sind alle die resonanzartigen Änderungen der HF-Eigenschaften der Probe, wo die Abmessungen der Probe wesentlich für den Resonanzeffekt sind. Dazu gehören alle Größeneffekte, die frequenzunabhängig sind, sowie die Fabry-Perot-Resonanzen von Magnetoplasmawellen, bei denen sich in einer Parallelplatte bei passender Wellenlänge $\lambda = \lambda(\omega_0, \boldsymbol{B}_0)$ durch Reflexion an den Deckflächen stehende Wellen bilden, die die HF-Eigenschaften der Probe stark beeinflussen (vgl. Abb. 6.9.0–1).

6.9.4.1. Magnetische Resonanzen. Zu diesen Resonanzphänomenen, die alle auf der Übereinstimmung der Anregungsfrequenz ω_0 mit der Präzessionsfrequenz Ω (vgl. 6.9.1.1.) magnetischer Momente in einem starken Magnetfeld beruhen, gehören die Kernresonanz (NMR), die paramagnetische Elektronenresonanz (EPR), die Quadrupolresonanz und die Resonanzen in Ferro- und Antiferromagnetika.

Aussagen über den Festkörper können dadurch gewonnen werden, daß die Beeinflussung der Energieeigenwerte des betrachteten Kern- bzw. Elektronenspins, deren Differenz proportional der Präzessionsfrequenz ist, durch die Felder untersucht wird, die ihre Ursache in der Struktur des Festkörpers haben (Lage der Spinmomente zueinander, Kristallfelder- und -störungen). Zum Nachweis von magnetischen Resonanzen befindet sich die Probe in einem statischen Magnetfeld $\boldsymbol{B}_0$. Die Anregung erfolgt über die magnetische Komponente $\boldsymbol{B}^i$ des HF-Feldes. Meßgröße ist die komplexe HF-Permeabilität μ^*, die das durch die Spinmomente mitbestimmte magnetische Verhalten des Festkörpers beschreibt.

Kernresonanz ist an den Spin der Atomkerne I geknüpft, der über das gyromagnetische Verhältnis γ mit dem magnetischen Kernmoment μ_K verbunden ist, $\mu_K = \gamma \hbar I$. Die untersuchten Kerne können Atome oder Isotope des den Kristall bildenden Stoffes bzw. Beimengungen genügender Konzentration sein.

Im quantenmechanischen Bild besitzt jeder Spin im Magnetfeld $\boldsymbol{B}_0$ $(2I+1)$ Einstellungsmöglichkeiten mit den dazu gehörigen Energieniveaus. Im einfachsten Fall ist $I = 1/2$, und der Spin kann sich parallel bzw. antiparallel zum Magnetfeld einstellen. Die beiden Energieniveaus sind um den Betrag $U = 2\mu_K B_0$ getrennt (vgl. Abb. 6.9.1–1). Im Resonanzfall ist die Photonenenergie gleich U, was der Bedingung $\omega_0 = \gamma B_0$ entspricht. Durch Anregen (Umklappen) der Kernspins findet eine Umbesetzung zwischen dem unteren und dem oberen Energieniveau statt, wodurch sich die Magnetisierung ändert. Angeregte Kerne können ihre Energie durch Stöße an andere Teilsysteme des Festkörpers abgeben und fallen in das untere Energieniveau zurück, so daß die Resonanz mit einer Absorption von HF-Energie verbunden ist. Die Effektivität dieser Prozesse läßt sich durch die *longitudinale Relaxationszeit* T_1 charakterisieren, die angibt, mit welcher Zeitkonstante das durch das HF-Feld angeregte Spinsystem nach Abschalten der Störung in das thermodynamische Gleichgewicht zurückkehrt.

Die magnetfeldabhängige Änderung der HF-Absorption in der Nähe der Resonanz ist die *Linie.* Die Linienbreite wird sowohl von T_1 als auch von der *transversalen Relaxationszeit* T_2 bestimmt. In klassischer Betrachtungsweise richtet das HF-Feld die rotierenden Anteile der magnetischen Dipolkomponenten der einzelnen Kerne phasenmäßig aus, wodurch ein oszillierendes magnetisches HF-Moment senkrecht zu $\boldsymbol{B}_0$ in der Probe entsteht. Der Zerfall dieser Phasenbeziehungen nach Abschalten des HF-Feldes wird durch T_2 charakterisiert. Er erfolgt im wesentlichen durch die Wechselwirkung der Spins untereinander.

Meßtechnisch lassen sich sowohl die HF-Absorption als auch das senkrecht zu $\boldsymbol{B}_0$ rotierende HF-Moment erfassen. Die Resonanzfrequenzen liegen im Radiofrequenzbereich, maximale Anregung erfolgt, wenn $\boldsymbol{B}^i$ senkrecht $\boldsymbol{B}_0$ ist.

In der HF-Spektroskopie des Festkörpers dient die Kernresonanz (NMR) als Methode, mit der man über theoretische Modelle die Eigenschaften des Kristalles untersuchen kann, die sich auf die Energieniveaus der Kernspins im Magnetfeld auswirken. Dabei führen:

a) Leitungselektronen zu einer Verschiebung der Resonanzfrequenz (Knight-shift),
b) innere Magnetfelder in Ferro- und Antiferromagnetika zu einer Beeinflussung der Kernresonanz,
c) Gitterdefekte zu einer Verbreiterung der Linie,
d) Wechselwirkungen zwischen den magnetischen Dipolen eines Kristallbausteines (Molekül) zu Linienaufspaltungen,
e) Phononen u. a. periodische Anregungen im Festkörper zu einer Beeinflussung der Relaxationszeiten.

Kernquadrupolresonanz. Kerne mit einem Spin größer 1/2 besitzen neben einem magnetischen Dipolmoment auch noch ein elektrisches Quadrupolmoment, auf das durch starke elektrische Feldgradienten Drehmomente ausgeübt werden können. Aus Symmetriegründen liegt das Drehmoment senkrecht zum Spin, so daß ohne Magnetfeld eine Präzessionsbewegung entsteht, wodurch das Kerndipolmoment eine oszillierende Komponente erhält, die mit dem HF-Feld koppelt. Bedingung dafür sind in Kernnähe starke (atomare) Feldgradienten. In Kristallen mit kovalent gebundenen Atomen können mit dieser Methode Aussagen über die Bindungsverhältnisse sowie die Struktur der Elementarzelle erhalten werden. In Ionenkristallen sind es vor allem die Feldgradienten von Gitterdefekten, die einer Messung zugänglich werden.

In einem Magnetfeld $\boldsymbol{B}_0$ führt die Quadrupolresonanz zur Aufspaltung und Dämpfung der Kernresonanz.

Paramagnetische Elektronenresonanz (EPR) beruht auf dem gleichen Wechselwirkungsmechanismus wie die Kernresonanz. Das gyromagnetische Verhältnis der Elektronen ist ≈ 2000mal größer als das der Kerne, weshalb diese Effekte vor allem im Mikrowellenbereich gemessen werden. EPR tritt dann auf, wenn die Summe aller zu einem Kristallbaustein gehörenden Elektronenspins nicht Null ist, z. B. in einer nichtabgesättigten Elektronenschale eines Atomes oder einem teilweise besetzten Band. Leitungselektronen zeigen jedoch nur geringe Resonanzeffekte, da für sie T_1 relativ klein ist. Untersucht werden vor allem Ionen von Übergangsmetallen mit einem von Null abweichenden Gesamtspin der Hüllenelektronen und z. B. auf Gitterfehlern lokalisierte Elektronen in Halbleitern.

Eine große Rolle spielt die Hyperfeinwechselwirkung mit den magnetischen Kerndipolen, in deren Nähe sich die Elektronen aufhalten. Elektronen haben wegen ihres Spins 1/2 zwei Energieniveaus, die dann entsprechend der Ausrichtung des Kernspins bezüglich $\boldsymbol{B}_0$ um kleine Energiebeträge verschoben werden, also aufspalten.

In Halbleitern lassen sich aus EPR-Untersuchungen Aussagen über Störstellen, Kristalldefekte und Oberflächenzustände gewinnen.

6.9.4.2. Zyklotronresonanz. Unter Zyklotronresonanz werden alle Resonanzphänomene zusammengefaßt, bei denen Ladungsträger, die sich um ein starkes Magnetfeld $\boldsymbol{B}_0$ auf

geschlossenen Umläufen oder Schraubenbahnen bewegen, Besonderheiten der HF-Eigenschaften der Probe hervorrufen, wenn sie durch die $\boldsymbol{E}^{\mathrm{i}}$-Komponente des HF-Feldes mit der Anregungsfrequenz ω_0 so beschleunigt werden, daß bei einem *Resonanzmagnetfeld* $\boldsymbol{B}_0^{\mathrm{r}}$ bei jedem Umlauf eine Energieerhöhung erfolgt.

Im Impuls- bzw. k-Raum (vgl. 6.3.1.), in dem das Bänderspektrum in Form von Isoenergieflächen darstellbar ist, läuft das Teilchen (Phasenpunkt) wegen der geringfügigen Energieerhöhung durch das HF-Feld allein durch die Wirkung von $\boldsymbol{B}_0$ auf einer ebenen Bahn $p_0 =$ konst. (vgl. 6.9.1.1.) auf einer Isoenergiefläche $E(\boldsymbol{k}) =$ konst. um. Wegen der Proportionalität zwischen Impuls $\boldsymbol{p}$ und $\boldsymbol{k}$-Vektor des Ladungsträgers gilt $p_0 = \hbar k_0$, wobei k_0, die Projektion des $\boldsymbol{k}$-Vektors auf die Magnetfeldrichtung, konstant ist und die Lage der Bahn auf der Isoenergiefläche bestimmt (vgl. Abb. 6.9.1–2). Die dabei umschriebene Fläche $S(E, k_0)$ liefert über $m^{\mathrm{c}}(k_0) = (\hbar^2/2\pi)\,\mathrm{d}S(k_0, E)/\mathrm{d}E$ den allgemeinsten Ausdruck für die Zyklotronmasse m^{c}. Sie bestimmt über die Beziehung $\Omega = eB_0/m^{\mathrm{c}}$ die Umlauffrequenz im Orts- und $\boldsymbol{k}$-Raum. Für einen parabolischen Bandverlauf $E(\boldsymbol{k}) = \sum (\hbar^2/2m_\beta)k_\beta^2$ sind die Isoenergieflächen Ellipsoide im $\boldsymbol{k}$-Raum. Die Zyklotronmasse ist dann für eine gegebene Magnetfeldrichtung konstant. Für $\boldsymbol{B}_0$ parallel zu einer Hauptachse der Ellipsoide gilt $m_\alpha^{\mathrm{c}} = \sqrt{m_\beta m_\gamma}$ (α, β, γ = 1, 2, 3 und zyklisch vertauscht).

Für alle Zyklotronresonanzeffekte ist die Anregungsfrequenz ω_0 gleich oder von gleicher Größenordnung wie Ω. Die Streuzeit τ muß dabei so groß sein, daß der Ladungsträger mehrere Umläufe zwischen zwei Stößen ausführt ($\Omega\tau \gg 1$). Unter Einsatz tiefer Temperaturen zur Unterdrükkung der Phononenstreuung ($T < 10\,\mathrm{K}$) und der Verwendung perfekter Einkristalle können Stoßzeiten im Bereich 10^{-11} bis 10^{-9} s erreicht werden, so daß für derartige Untersuchungen nur Mikro- oder Submillimeterwellen in Frage kommen.

Diamagnetische Resonanz. Für geringe Ladungsträgerkonzentration sind $\omega_0\tau \approx \Omega\tau \gg 1$ für Meßfrequenzen erfüllt, für die $\omega_0 > \omega_{\mathrm{p}}$ ist (vgl. Gl. 6.9.2–5). Unter lokalen Bedingungen ist dann λ relativ groß, und die Probendicke läßt sich kleiner als λ wählen (räumlich konstantes HF-Feld). $\boldsymbol{E}^{\mathrm{i}}$ verursacht eine HF-Polarisation $\boldsymbol{P}$ in der Probe, die sich aus der Polarisation des Untergrundes (vgl. 6.9.1.3.) und der durch $\boldsymbol{B}_0$ beeinflußten oszillatorischen Verschiebung der freibeweglichen Ladungsträger zusammensetzt.

$$\boldsymbol{P} = (\varepsilon_{\mathrm{L}} - \varepsilon_0)\,\boldsymbol{E}^{\mathrm{i}} + ne\boldsymbol{r}(t) = \left[(\varepsilon_{\mathrm{L}} - \varepsilon_0) + \frac{1}{\mathrm{i}\,\omega}\hat{\sigma}\right]\boldsymbol{E}^{\mathrm{i}},$$

($\boldsymbol{r}(t)$ = Auslenkung aus der kräftefreien Bahn).

Die absorbierte Leistung ist $W \sim \mathrm{Re}(\boldsymbol{E}^{\mathrm{i}}\,\mathrm{d}\boldsymbol{P}/\mathrm{d}t)$. Sie erhält Resonanzcharakter, wenn $\boldsymbol{E}^{\mathrm{i}}$ senkrecht zu $\boldsymbol{B}_0$ (zirkular) polarisiert ist. Legt man $\boldsymbol{E}^{\mathrm{i}}$ parallel zur y-Achse, wird $\mathrm{d}P/\mathrm{d}t \sim \mathrm{i}\omega_0 P_y \sim \sigma_{yy} E^{\mathrm{i}}$ und $W \sim \mathrm{Re}\,(\sigma_{yy})\,E^{\mathrm{i}2}$. Mit $\boldsymbol{B}_0$ parallel zur x-Achse besitzt σ_{yy} für $\omega_0 = \Omega = eB_0^{\mathrm{r}}/m^{\mathrm{c}}$ ein Maximum, sofern $\omega_0\tau$, $\Omega\tau \gg 1$ erfüllt ist (vgl. Gl. 6.9.3–1). Damit wird $m^{\mathrm{c}} = eB_0^{\mathrm{r}}/\omega_0$.

Für Proben mit endlichen Abmessungen, die sich isoliert im HF-Feld befinden, fällt wegen des Aufbaues eines inneren Gegenfeldes das Maximum der Energieabsorption nur für $\omega_{\mathrm{p}} \ll \omega_0$ mit der Zyklotronresonanz zusammen.

Meßgrößen. An reinen, optisch angeregten Halbleitern (Germanium, Silicium) werden bei tiefen Temperaturen Größe, Anzahl und Abhängigkeit der Zyklotronmassen vom Winkel des Magnetfeldes gegenüber den Kristallachsen bestimmt. Daraus werden über entsprechende Bandmodelle Aussagen über die Form der Isoenergieflächen in den relevanten Bandextrema erhalten. Aus dem Verlauf der HF-Absorption als Funktion von $\boldsymbol{B}_0$ und ω_0 (Linienform) wird die Streuzeit bestimmt. Aus dem Drehsinn der Welle bei zirkularer Anregung kann der Ladungsträgertyp ermittelt werden.

Zyklotronresonanz unter normalen Skineffekt-Bedingungen. Ist die Ladungsträgerkonzentration so hoch, daß im betrachteten Frequenzbereich $\omega_{\mathrm{p}} > \omega_0$ ist und damit die Skintiefe sehr klein wird (vgl. Gl. 6.9.2–5), wird es aus technologischen Gründen unmöglich, die Probendicke kleiner als diese zu halten.

Unter Einfluß von $\boldsymbol{B}_0$ kann HF-Energie in Form einer gedämpften Magnetoplasmawelle in die Probe eindringen. Sie wird durch die Wechselwirkung mit dem gesamten Ladungsträgerplasma bestimmt. Der einzelne Ladungsträger auf seiner Schraubenbahn ($\Omega\tau \gg 1$) in der Skinschicht wird durch das $\boldsymbol{E}^{\mathrm{i}}$-Feld der Welle beschleunigt. Resonanz tritt auf, wenn das $\boldsymbol{E}^{\mathrm{i}}$-Feld in der Lage ist, einer Ladungsträgergruppe bei jedem Umlauf einen Energiezuwachs zu vermitteln. Von Bedeutung ist das für die Untersuchung von stark dotierten Halbleitern, bei denen die Konzentration einer Ladungsträgersorte überwiegt ($n \neq p$, unkompensiertes Plasma).

In unkompensierten Plasmen mit $\boldsymbol{B}_0$ parallel zur Oberfläche der Probe existieren zwei Typen von Zyklotronresonanzen:

Tilted orbit Resonanz entsteht, wenn $\boldsymbol{E}(0)$ parallel $\boldsymbol{B}_0$ gewählt wurde und Ladungsträger in der Skinschicht Schraubenbahnen mit stark vom Rotationswinkel abhängenden Anstieg durchlaufen (vgl. Abb. 6.9.1–2). Die zugehörige Magnetoplasmawelle (ordentliche Welle) ist linear parallel zu $\boldsymbol{B}_0$ polarisiert. Ihr $\boldsymbol{E}^{\mathrm{i}}$ beschleunigt jene Ladungsträger, die große zu ihr parallel verlaufende Bahnanteile besitzen.

Im Falle eines parabolischen Bandverlaufes entstehen derartige Bahnen, bei großer Anisotropie der ellipsoidförmigen Isoenergieflächen, wenn $\boldsymbol{B}_0$ „schief" zu den Hauptachsen liegt. Für $\omega_0 = \Omega$, der Umlauffrequenz auf der Schraubenbahn, besitzt die Oberflächenimpedanz Resonanzcharakter.

Plasmazyklotronresonanz tritt in Festkörpern auf, die mehrere Ladungsträgergruppen eines Typs mit stark abweichender Anisotropie der effektiven Massen besitzen, und wenn $\boldsymbol{E}(0)$ senkrecht $\boldsymbol{B}_0$ gewählt wurde. Die zugehörige Magnetoplasmawelle (außerordentliche Welle) ist eine Kombination von Raumladungs- und Transversalwelle, bei der der elektrische Feldstärkevektor $\boldsymbol{E}^{\mathrm{i}}$ entgegengesetzt zu den Zyklotronumläufen um $\boldsymbol{B}_0$ rotiert. Im Resonanzfall für eine Teilchengruppe $\omega_0 = \Omega$ erfolgt trotz entgegengesetzten Drehsinns eine Energieerhöhung, da wegen der großen Bahnexzentrizität Abbremsung und Beschleunigung durch $\boldsymbol{E}^{\mathrm{i}}$ sich nicht aufheben.

Resonanz in Faraday-Konfiguration. Zeigt das Magnetfeld senkrecht zur Oberfläche, ist in unkompensierten Plasmen die Zyklotronresonanz die Einsatzkante der Helikonen (vgl. 6.9.3.1.). Das führt zu einem verstärkten Energietransport in die Probe, wenn $\Omega \gtrsim \omega_0$ bzw. $\boldsymbol{B}_0 \gtrsim \boldsymbol{B}_0^{\mathrm{r}}$ wird. Die Oberflächenimpedanz als Funktion des Magnetfeldes bei vorgegebener Anregungsfrequenz hat in der Nähe der Resonanzstelle monotonen ansteigenden Verlauf, ihre Ableitung nach dem Magnetfeld besitzt bei $\omega_0 = \Omega$ ein Extremum.

Azbel-Kaner-Zyklotronresonanz. In reinen Metallen bei tiefen Temperaturen (nichtlokale Bedingungen) treten Resonanzen in der Oberflächenimpedanz auf, wenn das Ma-

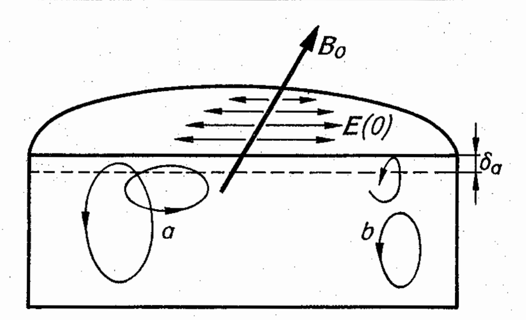


Abb. 6.9.4–1 Projektion von Umlaufbahnen auf eine Schnittebene der Probe unter nichtlokalen Bedingungen, (a) effektive, (b) ineffektive Bahnen (δ_a = anomale Skintiefe, $\boldsymbol{E}(0)$ = $\boldsymbol{E}$-Vektoren des Anregungsfeldes auf der Oberfläche)

gnetfeld $\boldsymbol{B}_0$ streng parallel zur Oberfläche liegt. Die anomale Skinschicht entsteht unter der Wirkung des Magnetfeldes durch die Ladungsträger, die sich dicht unter der Oberfläche fast parallel zu ihr bewegen (Bahnscheitelpunkt). Der Zyklotronradius ist dabei viel größer als die Skintiefe δ_a. Diese Ladungsträger kehren mehrfach ($\Omega\tau \gg 1$) auf ihrer Zyklotronbahn in die Skinschicht zurück. Finden sie nach dem Umlauf das elektrische Feld in der gleichen Richtung vor, erfolgt ein Aufschaukeln der Energie, ähnlich wie im Spalt eines Zyklotrons. Das Teilchen wird nur kurzzeitig beschleunigt und fliegt auf dem größten Teil der Bahn im feldfreien Raum (Abb. 6.9.4–1).

Für eine feste Anregungsfrequenz ω_0 findet damit auch eine Resonanz statt, wenn das Teilchen in schwächeren Magnetfeldern nach mehreren Perioden des HF-Feldes die Skinschicht erreicht (Harmonische, $\omega_0 = n\Omega$, n = ganzzahlig).

In Metallen tragen die Ladungsträger zum Strom bei, deren Energie gleich der Fermi-Energie ist. Bestimmend für das Verhalten ist die Isoenergiefläche mit der Fermi-Energie, die *Fermi-Fläche.* Besonderheiten in der Oberflächenimpedanz liefern nur Zyklotronbahnen, die speziellen Umläufen auf der Fermi-Fläche entsprechen, wo die Umlauffrequenz wenig von k_0 abhängt.

Extremalbahnen sind Umläufe, für die $\Omega(k_0)$ ein Extremum besitzt (Abb. 6.9.4–2). Im Normalfall ist dort auch die umschriebene Fläche S extremal. Für $\boldsymbol{E}(0)$ senkrecht $\boldsymbol{B}_0$ erfolgt eine maximale Anregung.

Grenzpunktbahnen. An den Punkten, wo die Oberflächennormale auf die Fermi-Fläche parallel zu $\boldsymbol{B}_0$ ist, läßt sich die Umgebung in Form eines Ellipsoids approximieren (m^c unabhängig von k_0) (vgl. Abb. 6.9.4–2). Diese Teilchen durchlaufen im Ortsraum langgestreckte Schraubenbahnen mit einheitlicher Umlauffrequenz. Für $\boldsymbol{E}(0)$ parallel $\boldsymbol{B}_0$ liefern sie eine Resonanz der Oberflächenimpedanz.

Meßgrößen bei der Azbel-Kaner-Resonanz sind die Zyklotronmassen von Extremal- und Grenzpunktbahnen, die als Funktion des Winkels des Magnetfeldes gegenüber den Kristallachsen bestimmt werden. Aufgrund der komplizierten Struktur der Fermi-Fläche der meisten Metalle liefern die Extremalbahnen in erster Linie Aussagen über deren Form und die Grenzpunktmassen Aussagen über deren Krümmung im Grenzpunkt.

6.9.4.3. Oberflächeneffekte. In Metallen und Halbleitern können durch den Einfluß der Oberfläche teilweise im Zusammenwirken mit einem elektrischen oder magnetischen

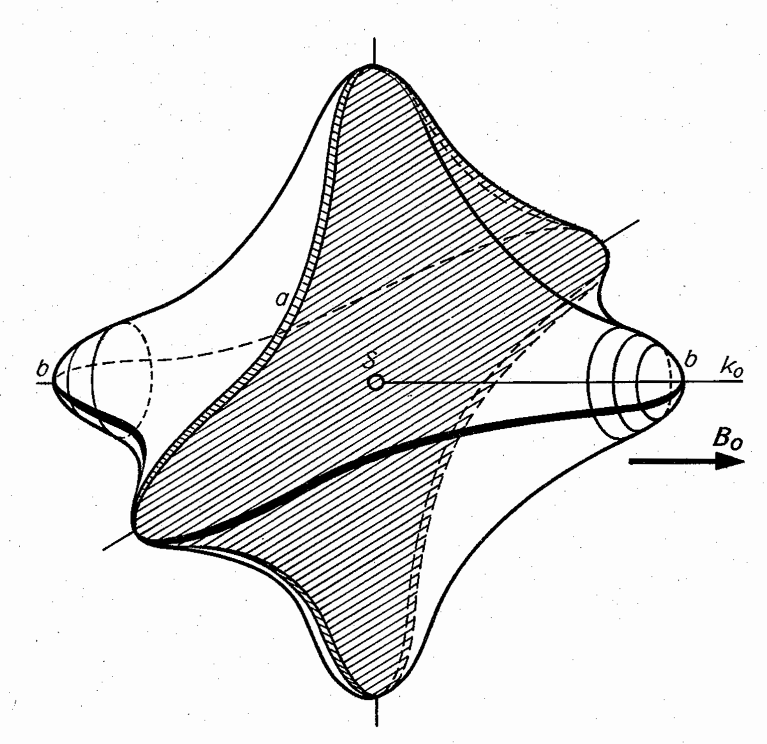


Abb. 6.9.4–2 Bahntypen auf einer Fermi-Fläche, (a) Extremalbahn, (b) Grenzpunktbahn

Feld neue Quantenzustände entstehen, zwischen denen durch ein HF-Feld Übergänge angeregt werden.

Subbänder. Erzeugt man durch Feldeffekt oder Oberflächenladungen in einem Halbleiter dicht unter der Oberfläche ein Gebiet, in dem sich Ladungsträger anderer Polarität als im Volumen aufhalten (Inversionsgebiet), so befinden sich diese in einem Potentialtopf. Die Teilchenenergie wird quantisiert. Parallel zur Oberfläche behalten sie die Eigenschaften freier Teilchen. Dadurch entstehen Subbänder. Mit den Methoden der Submillimeterspektroskopie konnte in Silicium, Germanium und Indium-Antimonid das Verhalten der Subbänder untersucht werden.

Magnetische Oberflächenzustände. In Metallen unter stark nichtlokalen Bedingungen ($\delta_a \ll l$) tritt in schwachen Magnetfeldern ($\boldsymbol{B}_0$ parallel zur Oberfläche) für spiegelnd an der Oberfläche reflektierte Ladungsträger eine zusätzliche Energiequantelung auf. Die Ladungsträger vollführen periodische Girlandenbahnen mit einem sehr großen Krümmungsradius R_c. Aus allgemeinen Prinzipien ist abzuleiten, daß der von einem Bogen und der Oberfläche eingeschlossene Magnetfluß $F\boldsymbol{B}_0 = \Phi$ (F = Segmentfläche), eine ganze Zahl n plus Phasenfaktor γ von Flußquanten Φ_0 ist: $FB_0 = \Phi_0(n+\gamma) = (4/3)\sqrt{2x_n^3 R_c}\, B_0 = (h/e)(n+\gamma)$. Damit ist die Eindringtiefe der Bahn x_n quantisiert. Die zugehörigen Energieeigenwerte sind $E_n = eB_0 v x_n$ (v = Geschwindigkeitskomponente der Ladungsträger parallel zur Oberfläche, senkrecht $\boldsymbol{B}_0$) (Abb. 6.9.4–3). Das HF-Feld erzeugt Übergänge zwischen diesen Niveaus, wenn die Photonenenergie der Differenz zwischen Ausgangs- (m) und Endniveau (n) entspricht, was ähnlich wie beim Wasserstoffspektrum Serien ergibt.

Ein Girlandenbogen ist ein Teilstück eines Zyklotronumlaufes, eines Phasenpunkts auf der Fermi-Fläche mit dem Krümmungsradius k_r (Achse parallel $\boldsymbol{B}_0$), wo die Fermi-Geschwindigkeit $\boldsymbol{v}$ (parallel zur Normalen auf die Fermi-Fläche) etwa auch zur Oberfläche parallel ist. k_r ist proportional R_c. Signifikante Änderungen der Oberflächenimpedanz ergeben sich nur, wenn viele Ladungsträger in k_r und v übereinstimmen, sich also ein großer Teil der Fermi-Fläche durch ein Zylindersegment approximieren läßt.

Meßgrößen sind k_r und die Streuzeit τ auf Teilstücken der Fermi-Fläche.

6.9.4.4. Größeneffekte. Größeneffekte treten nur in hochreinen Metallen bei tiefen Temperaturen auf, wenn die freie Weglänge l größer als die Probendicke ist und anomale Skineffektbedingungen vorliegen. Legt man dann das Magnetfeld parallel zur Oberfläche einer planparallelen Probe, kann $\boldsymbol{B}_0$ soweit erniedrigt werden, daß der Durchmesser der Umlaufbahn von Ladungsträgern, die zu einer Extremalbahn gehören (vgl. 6.9.4.2.), gleich oder größer als die Probendicke wird.

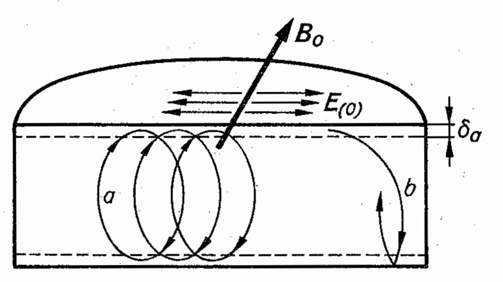


Abb. 6.9.4–4 Radiofrequenzgrößeneffekt, (a) Umlaufbahn in der Nähe der Resonanz, (b) gestreute Bahn (δ_a = anomale Skintiefe, $\boldsymbol{E}_{(0)}$ = $\boldsymbol{E}$-Vektoren des Anregungsfeldes auf der Oberfläche)

Abschneiden der Azbel-Kaner-Zyklotronresonanz. Wird der Durchmesser einer Umlaufbahn, der zu einer Harmonischen gehört, größer als die Probendicke, fallen ab dieser alle weiteren Harmonischen aus, was es erlaubt, den Durchmesser abzuschätzen.

Radiofrequenzgrößeneffekt. Darunter versteht man die Erscheinung, daß sich auch im Radiofrequenzbereich die Oberflächenimpedanz ändert, wenn durch Verkleinern von $\boldsymbol{B}_0$ eine bestimmte Extremalbahn unterhalb $\boldsymbol{B}_0^a$ nicht mehr in die Probe „paßt" (Abb. 6.9.4–4).

Meßgrößen. Die Projektion der Umlaufbahn im Ortsraum auf die Ebene senkrecht zum Magnetfeld $\boldsymbol{B}_0$ und der Umlauf auf der Fermi-Fläche lassen sich durch eine Drehung um 90° um $\boldsymbol{B}_0$ und eine Maßstabsänderung aufeinander abbilden. Damit ist es möglich, den Bahndurchmesser (entspricht Probendicke d) in den senkrecht dazu liegenden k-Vektor der Extremalbahn umzurechnen und durch Magnetfelddrehung Schnitte durch die Fermi-Fläche auszumessen, $k = (e/\hbar)\,\boldsymbol{B}_0^a d$.

6.10. Oberflächen

6.10.1. Allgemeine Charakterisierung von Oberflächen

Festkörper haben eine *Oberfläche*, oder allgemeiner *Grenzfläche*, an der sie mit Phasen anderer Struktur und Zusammensetzung in Berührung stehen. Besondere Bedeutung ge-

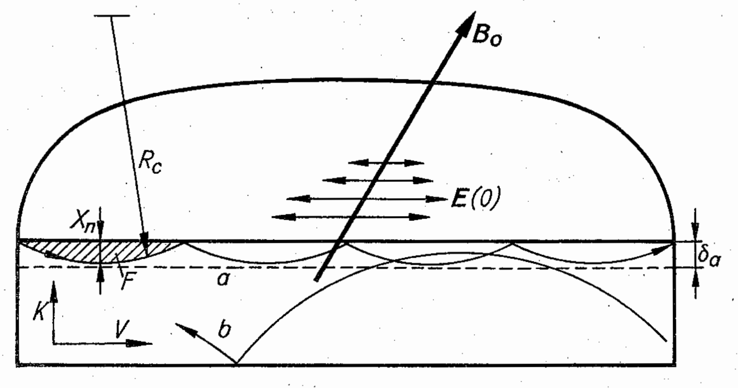


Abb. 6.9.4–3 Bahntypen unter stark nichtlokalen Bedingungen in schwachen Magnetfeldern, (a) Girlandenbahn (quantisiert), (b) Zyklotronbahn (abgeschnitten)

winnt die Festkörperoberfläche dadurch, daß eine Reihe wichtiger physikalischer und chemischer Prozesse, wie Phasenumwandlungen (Wachstum, Schmelzen, Auflösung, Verdampfung), mechanischer und chemischer Angriff (Verschleiß, Korrosion, Passivierung) und chemische Umsetzungen (Katalyse), sich an der Oberfläche abspielen und durch deren physikalisch-chemische Eigenschaften stark beeinflußt werden. Die durch die *Morphologie, Struktur, chemische Zusammensetzung* und *elektronische Struktur* bestimmten Eigenschaften der Festkörperoberfläche stellen den Untersuchungsgegenstand der *Oberflächenphysik* und *Oberflächenchemie* dar.

6.10.1.1. Technische Oberflächen. Ihrer Oberflächenmorphologie, Struktur und chemischen Zusammensetzung nach sind *technische Oberflächen* relativ inhomogen. Sie werden je nach Verwendungszweck durch mechanische Bearbeitung (Zerspanen, Schleifen, Polieren), chemische Bearbeitung (Ätzen, Polieren), chemische Reaktionen (Passivierung, Oxidation) oder durch unterschiedliche Schichtabscheidungsverfahren (Aufdampfung, Ionenstrahl-Abscheidung, chemische oder elektrolytische Schichtabscheidung) hergestellt. Die Morphologie technischer Oberflächen ist in Abhängigkeit von der Vorbehandlung durch eine Rauhtiefe von bestenfalls 10^{-1} µm charakterisiert. Die kristallografische Struktur wird von der polykristallinen Struktur des Volumenmaterials bestimmt, Schichtoberflächen sind meist feinkristallin und Reaktionsschichten oft amorph. Je nach Vorbehandlung enthalten technische Oberflächen große Anteile an Gitterstörungen, wie Mikrorisse, Versetzungen. Die chemische Zusammensetzung ist nicht homogen, z. B. durch in die Oberfläche eingearbeitete Schleifmittel, und hängt von Reaktionen mit Verunreinigungen und der Atmosphäre während der Bearbeitung ab.

6.10.1.2. Physikalische Oberflächen. Oberflächen, deren Morphologie, kristallografische Struktur, chemische Zusammensetzung und Elektronenstruktur sich durch physikalische Meßmethoden möglichst vollständig charakterisieren lassen, können als *physikalische Oberflächen* definiert werden. Eine besondere Bedeutung besitzen in der Grundlagenforschung und in modernen Technologien Kristalloberflächen, bei denen die strukturelle Besonderheit der Oberfläche, die nur zweidimensionale Periodizität parallel zur Oberfläche gegenüber der im Volumen gültigen dreidimensionalen Periodizität, besonders deutlich wird.

Der Schwerpunkt experimenteller Forschung liegt bei der Untersuchung der *Festkörper-Gas-Oberfläche*, da die meisten modernen Analysetechniken auf die Festkörper-Flüssigkeits-Grenzfläche wegen der vergleichbaren Dichten beider Phasen nicht anwendbar sind. Für die Herstellung und Analyse von physikalischen Oberflächen müssen eine Reihe von Voraussetzungen erfüllt sein, deren wichtigste die *Vakuumbedingungen*, geeignete *Präparations-* und *Reinigungsverfahren* sind. Wegen der unvermeidbaren Adsorptionsprozesse bei Anwesenheit von Gasmolekülen in der umgebenden Phase können Oberflächen in chemisch reinem, adsorptionsfreiem Zustand nur unter Ultrahochvakuumbedingungen unterhalb 10^{-7} Pa hergestellt und rein gehalten werden. Strukturell möglichst perfekte Kristalloberflächen werden an kompakten Proben durch röntgenografisches Orientieren, Schneiden und chemisches bzw. elektrolytisches Polieren präpariert. Einige Element- bzw. Ionenkristalle lassen sich nach niedrig indizierten Kristallorientierungen spalten; nur bei wenigen Substanzen gelingt die Präparation perfekter Wachstumsflächen aus der Lösung bzw. Dampfphase. In zunehmendem Maße wird die Dünnschichtabscheidung zur Herstellung von Einkristalloberflächen in Form von epitaktischen (kristallografisch gesetzmäßig verwachsenen) Schichten auf einkristallinen Substratunterlagen benutzt.

In den Meßkammern der Analysengeräte sollten ebenfalls Ultrahochvakuumbedingungen herrschen, damit während des Meßvorganges keine das Ergebnis beeinflussende Kontamination durch Adsorption aus dem Restgas auftreten kann. Als Reinigungsverfahren für außerhalb der Meßkammern präparierte Oberflächen werden angewendet: atomschichtweises Abtragen der Oberfläche durch Beschuß mit Edelgasionen *(Ionensputtern)* und anschließende Ausheilung der induzierten Kristalldefekte durch Erwärmung, Aufheizen der Probe auf für die Desorption bzw. Verdampfung von Verunreinigungen und arteigenem Material hinreichend hohe Temperaturen, chemische Oxidations- bzw. Reduktionsprozesse mit geeigneten Gasen zur Beseitigung spezifischer Oberflächenverunreinigungen.

6.10.2. Makroskopische Zustandseigenschaften

An der Oberfläche von Festkörpern befinden sich die atomaren oder molekularen Bausteine in einem anderen geometrischen und energetischen Zustand als im Volumen. Sie grenzen entweder an das Vakuum oder an Teilchen einer fremden Phase, so daß sich charakteristische Änderungen in der Wechselwirkung der Oberflächenbausteine untereinander und zu artfremden Bausteinen ergeben. Die Wechselwirkungskräfte bestimmen aber die makroskopischen Eigenschaften der Oberfläche, z. B. die makroskopische Form wachsender Kristalle und die Orientierung von Spaltflächen und Adsorptions- bzw. Segregationsprozesse.

6.10.2.1. Oberflächenenergie von Festkörpern. Bei der Diskussion der thermodynamischen Eigenschaften eines Systems spielen die freien Energien eine wichtige Rolle, da sie zur Formulierung von Gleichgewichtskriterien benutzt werden können. Eine ähnliche Bedeutung hat die *Oberflächenenergie* bei der thermodynamischen Behandlung von Oberflächeneigenschaften. Sie ist definiert als diejenige freie Energie (Enthalpie), die zur Vergrößerung der Oberfläche aufzuwenden ist und besitzt die Dimension einer Energie je Flächeneinheit. Die Oberflächenenergie kann auf mehrere Arten näherungsweise gemessen oder theoretisch abgeschätzt werden. So ist z. B. die bei der *Spaltung* von Kristallen je Flächeneinheit aufzuwendende Arbeit, abzüglich irreversibler Anteile für plastisches Fließen, Wärme, Punktdefekte, zur Messung der Oberflächenenergie geeignet. Die *Spaltbarkeitsebenen* fallen mit denjenigen Flächenorientierungen zusammen, für die die Oberflächenenergie kleinste Werte annimmt, z. B. [100]-Flächen beim NaCl-Kristall. Die *Anisotropie* (Orientierungsabhängigkeit) der Oberflächenenergie von Kristallen hat für das Kristallwachstum eine große Bedeutung und erklärt das Auftreten und die Stabilität nur weniger niedrig indizierter Oberflächenorientierungen.

Die *Gleichgewichtsform* eines Kristalls ist diejenige Form, die mit der umgebenden Phase, z. B. Dampf, im Gleichgewicht steht. Sie besitzt nach GIBBS-WULFF bei vorgegebenem Volumen die kleinste Oberflächenenergie.

Der geometrische Zusammenhang zwischen der Gleichgewichtsform G(hkl) und dem Polardiagramm der orientierungsabhängigen Oberflächenenergie σ(hkl), die sogenannte Wulff-Konstruktion, ist in Abb. 6.10.2–1 dargestellt. Die Richtung eines Radiusvektors steht senkrecht auf der

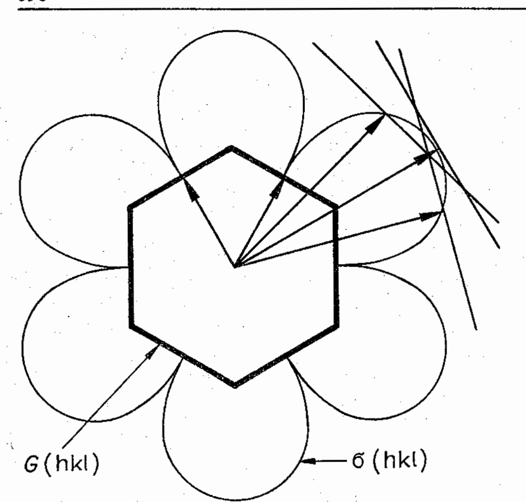


Abb. 6.10.2–1 Wulff-Konstruktion der Gleichgewichtsform G(hkl) aus dem Polardiagramm der Oberflächenenergie σ(hkl)

durch die *Millerschen Indizes* (hkl) gegebenen Flächenorientierung und die Länge des Vektors ist gleich der Oberflächenenergie für diese Orientierung. Die Gleichgewichtsform G(hkl) ergibt sich als innere Einhüllende aller Wulff-Ebenen, die senkrecht zu den Radiusvektoren an ihren Endpunkten gezeichnet werden. Die Gleichgewichtsformflächen können unter bestimmten Bedingungen, wie Abwesenheit von äußeren Kräften und Fremdadsorption, im Anfangsstadium des Wachstums an Kristallen als *Wachstumsflächen* beobachtet werden.

Die *Adsorption* aus der Gasphase, die in der Anreicherung von Atomen bzw. Molekülen an der Oberfläche unter Bildung von Mono- oder Multischichtbedeckungen besteht, ist nach GIBBS mit einer Erniedrigung der Oberflächenenergie des Festkörpers verbunden. In analoger Weise läßt sich auch die *Segregation* als Anreicherung eines im Festkörper gelösten Stoffes an der Oberfläche durch Verminderung der Oberflächenenergie erklären.

6.10.2.2. Morphologie physikalischer Oberflächen. Zu den *Abbildungsmethoden makroskopischer Strukturfehler*, wie Kratzer, hohe Spaltstufen, Einschlüsse, Facetten, gehören die normale Auflichtmikroskopie, die Interferenzmikroskopie, die Rasterelektronenmikroskopie und die indirekten Oberflächenabdruckverfahren für die Durchstrahlungselektronenmikroskopie. *Reale Oberflächen* haben einen unvermeidbaren Anteil an *mikroskopisch struktureller Fehlordnung* (vgl. Abb. 6.10.5–1), zu der Punktdefekte (Oberflächenlücken, adsorbierte Atome), Stufen, Versetzungen und Korngrenzen gehören. Die Ursachen dafür liegen einerseits in den Herstellungsbedingungen bzw. Reinigungsverfahren und andererseits stellt sich an jeder Oberfläche bei endlichen Temperaturen eine thermische Fehlordnung, wie Lücken, Adatome, Stufen, ein. Das wichtigste indirekte Abbildungsverfahren für mikroskopische Strukturfehler auf Kristalloberflächen ist das *Dekorationsverfahren*, das auf der bevorzugten Keimbildung von Schwermetallatomen (Au, Pt) oder Ionenkristallen (AgBr, CsI) an atomaren Stufen, Versetzungen und Punktdefekten gegenüber der Keimbildung auf störungsfreien Flächenbereichen beruht.

Für die *direkte Abbildung atomarer Stufen* auf Kristalloberflächen sind drei Methoden bekannt, die aber mit einer sehr speziellen Probengeometrie bzw. -präparation verbunden sind: Die Feldionenmikroskopie (vgl. 6.10.3.3., Taf. 22), die Reflexionselektronenmikroskopie bei streifendem Einfall unter Ultrahochvakuumbedingungen und die Transmissionselektronenmikroskopie dünner Schichten im Dickenbereich unterhalb 20 nm.

6.10.3. Mikroskopische Zustandseigenschaften.

Zum Verständnis der Elementarprozesse der an Oberflächen ablaufenden physikalischen und chemischen Vorgänge, wie Phasenumwandlungen, chemische Reaktionen, ist es erforderlich, die *fundamentalen Zustandseigenschaften* von Oberflächen, wie *Oberflächenstruktur, chemische Zusam-*

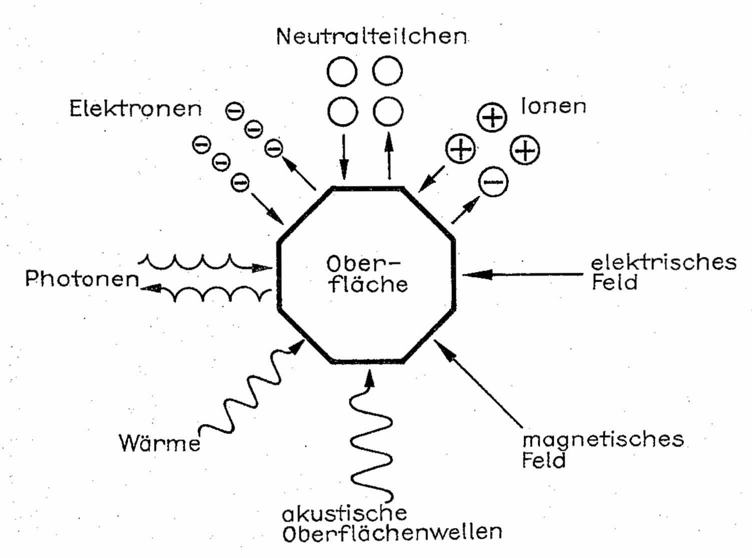


Abb. 6.10.3–1 Bildhafte Darstellung des Prinzipes der Oberflächenanalyse-Techniken

mensetzung, Elektronenstruktur und *Bewegung von Oberflächenatomen*, im mikroskopischen Maßstab zu messen.
Die bei der Wechselwirkung von Teilchen oder Feldern mit dem Festkörper emittierten bzw. gestreuten Teilchen tragen auch Informationen über die Oberfläche. Durch geeignete Kombination der in Abb. 6.10.3–1 dargestellten Anregungs- (Eingangspfeil) bzw. Emissionsprozesse (Ausgangspfeil) ergeben sich eine Vielzahl möglicher Analysemethoden. Die Informationsgewinnung aus den emittierten Teilchen erfolgt in speziellen Detektoren oder Analysatoren durch Identifizierung der Partikeln, Messung der Energieverteilung oder Messung der räumlichen Verteilung.

6.10.3.1. Oberflächenanalyse-Techniken. Für die Oberflächenanalytik eignen sich nur Prozesse starker Wechselwirkung der einfallenden bzw. emittierten Teilchen mit den Festkörperatomen, so daß die *Informationstiefe* auf ein Oberflächengebiet von wenigen Atomlagen beschränkt bleibt. Solche Prozesse sind z. B. die elastische und inelastische Streuung von langsamen bzw. mittelschnellen Elektronen an den Ionenrümpfen und Elektronen der Oberflächenatome, die binären Stoßprozesse zwischen niederenergetischen Ionen bzw. Neutralteilchen und Atomen und die Schwingungsanregung von Oberflächenatomen durch Wärme oder Felder. Neben der *Oberflächenempfindlichkeit* wird ein hohes *laterales Auflösungsvermögen* der Methoden angestrebt, um z. B. inhomogene Elementverteilungen nachzuweisen und unter Anwendung des Rasterprinzips ein entsprechendes *abbildendes Analyseverfahren* zu realisieren. Tab. 6.10.3–2 zeigt das Anregungs-Emissionsschema für die wichtigsten oberflächenempfindlichen Analysetechniken. Die Meßmethoden lassen sich den zu messenden fundamentalen Oberflächeneigenschaften wie folgt zuordnen: Oberflächenstruktur – LEED, RHEED, FIM, ISS (vgl. 6.10.3.3.), Element- bzw. chemische Analyse – AES, XPS, SIMS, ISS (vgl. 6.10.3.2.), Elektronenstruktur – ELS, XPS, INS, FES (vgl. 6.10.4.), Atombewegungen – FIM (vgl. 6.10.3.3.).

6.10.3.2. Methoden zur Element- bzw. chemischen Analyse von Oberflächen. AES – Augerelektronenspektroskopie. Durch Primärelektronen werden innere Elektronenschalen eines Atoms ionisiert und die nachfolgende Relaxation des angeregten Zustandes verläuft über strahlungslose Augerprozesse, die zur Emission von Elektronen führen, deren Energien charakteristisch für das Element und die speziellen Elektronenübergänge sind und sich somit zur Elementanalyse eignen.

XPS – Fotoelektronenspektroskopie. Röntgen- bzw. UV-Strahlung löst durch Fotoeffekt Elektronen aus inneren Elektronenschalen bzw. Valenzbändern, deren Energie- und Winkelverteilung in hochauflösenden Spektrometern gemessen wird und die die Bestimmung der Elementzusammensetzung, des chemischen Bindungszustandes und der Elektronenstruktur gestatten.

SIMS – Sekundärionen-Massenspektroskopie. Durch Primärionenbeschuß direkt an der Oberfläche abgelöste positive und negative Ionen bzw. Molekülionen werden in Massenspektrometern identifiziert und liefern eine chemische Analyse sehr hoher Nachweisempfindlichkeit.

ISS – Ionenstreuungsspektroskopie. Die Messung der Energieverteilung elastisch rückgestreuter Ionen erlaubt im Fall niederenergetischer Primärionen eine Elementanalyse nur der obersten Atomschicht, bei hochenergetischen Primärionen (Rutherford-Rückstreuung) reicht der analysierte Bereich ins Volumen hinein. Tab. 6.10.3–3 gibt eine Zusammenstellung der wichtigsten vergleichbaren Leistungsparameter der Analysemethoden.

Typische Meßaufgaben für die chemische Oberflächenanalytik sind: Reinheits- bzw. Verunreinigungsanalyse vor einem Experiment zur Untersuchung von Oberflächeneigenschaften oder zwischen Prozeßstufen moderner industrieller Technologien (Mikroelektronik), Oberflächenkonzentrationsmessungen bei Adsorptions-, chemischen Reaktions-, Diffusions- und Segregationsprozessen, Konzentrationstiefenprofilmessungen in der Oberflächenschicht durch Kombination einer Analysemethode mit dem Ionensputter-Verfahren (vgl. 6.10.1.2.).

Tab. 6.10.3-2 Anregungs-Emissionsschema für Oberflächenanalysemethoden

Anregung / Emission	Photon	Elektron	Ion	elektrisches Feld	Neutralteilchen
Elektron	XPS UPS	LEED RHEED AES ELS	INS	FES	
Ion			SIMS ISS	FIM	
Neutralteilchen					Atomstreuung

6.10.3.3. Atomanordnung an Kristalloberflächen. Oberflächenkristallografie. In vielen Fällen trifft die Annahme zu, daß die Oberflächenatome die Gitterplätze beibehalten, die auch im Inneren des Kristalls eingenommen werden. Dann läßt sich die Atomanordnung an der Oberfläche leicht durch die Kristallstruktur des betreffenden Materials und durch die die Oberfläche kennzeichnenden *Millerschen Indizes* (hkl) angeben (vgl. 6.1.2.2.). Die zur Oberfläche parallelen Netzebenen haben alle die gleiche zweidimensionale periodische Anordnung der Atome, die am dichtesten in kristallografischen Netzebenen mit kleinen Millerschen Indizes gepackt sind. Oberflächen mit hohen Millerschen Indizes bilden eine Terrassenstruktur, wobei die Terrassen

Tab. 6.10.3–3 Leistungsparameter der Methoden zur Elementaranalyse und chemischen Analyse von Oberflächen

	XPS	AES	SIMS	ISS
Informationstiefe	1...2,5 nm	0,4...2,5 nm	0,3...2 nm	0,3...10^3 nm
Elementnachweis	≧ He	≧ Li	≧ H	> H
Nachweisgrenze	1 Atom-%	0,1 Atom-%	10^{-5} Atom-%	0,1...1 Atom-%
chemischer Bindungszustand	ja	(ja)	ja	nein
laterale Auflösung	mm	0,1 µm	1 µm	mm
Abbildungsverfahren	nein	ja	ja	nein

wieder dichtgepackte Netzebenen darstellen. Enthält die Oberflächennetzebene eines Kristallgitters, das sich aus zwei Sorten von Gitterbausteinen zusammensetzt, nur eine Atomsorte, so spricht man von *polaren Oberflächen.* Beispiele dafür sind die (111)-Oberflächen des Zinkblende- und NaCl-Gitters oder die (0001)-Oberfläche des Wurtzitgitters. Elektrostatisch stabil sind jedoch nur die *nichtpolaren Oberflächen,* die eine gleichmäßige Zusammensetzung beider Gitterbausteine aufweisen, z. B. die NaCl-(100)-, Zinkblende-$(1\bar{1}0)$-, Wurtzit-$(11\bar{2}0)$- und -$(10\bar{1}0)$-Oberflächen.

Nachweismethoden zur Atomanordnung. Die allgemeine Methode zur Bestimmung der Atompositionen in einer periodischen Anordnung mit dem interatomaren Abstand a ist die Beugung einfallender Strahlung, deren Wellenlänge $\lambda < 2a$ sein muß. Für Atomgitter bietet sich neben der Röntgen-Strahlung die Partikelstrahlung an, deren Wellenlänge sich nach der De-Broglie-Beziehung $\lambda = h/\sqrt{2mE}$ bestimmen läßt (vgl. 4.1.3.). Für Oberflächenuntersuchungen eignen sich jedoch nur Elektronen und leichte Atome bzw. Ionen, da Röntgen-Strahlen und Neutronen wegen ihrer schwachen Wechselwirkung mit den Atomen tief in den Kristall eindringen und nicht oberflächenempfindlich sind. Für Metalle mit hohem Schmelzpunkt ermöglicht die *Feldionenmikroskopie* im Gegensatz zu den Beugungsmethoden auch die direkte Abbildung der Atompositionen.

Elektronenbeugung. Die von der Energie abhängige Eindringtiefe der Elektronen führt zu zwei unterschiedlichen Beugungsanordnungen für Oberflächenstrukturuntersuchungen. Niederenergetische Elektronen (20 bis 500 eV, *LEED = Low-Energy Electron Diffraction*) können senkrecht, höherenergetische Elektronen (5 bis 50 keV, *RHEED = Reflection High-Energy Electron Diffraction*) müssen streifend auf die Oberfläche auftreffen, damit die zur Oberfläche senkrechte Eindringtiefe auf wenige Netzebenen beschränkt bleibt. Die vollständige Bestimmung der Atomanordnung im Oberflächengebiet ist aus dem Beugungsbild allein nicht möglich, sondern erfordert die Intensitätsanalyse der einzelnen Reflexe.

Atom- und Ionenstreuung. Die Beugung von leichten Atomen (H, D_2, He, Ne, Ar), obwohl experimentell schon 1930 durch ESTERMANN und STERN nachgewiesen, gelang aufgrund von Schwierigkeiten bei der Erzeugung monoenergetischer Strahlen und ihrer Analyse nach dem Streuprozeß bisher nur an Alkalihalogeniden und einigen Metalloberflächen. Eine breitere Anwendung finden daher Ionenstreuexperimente, bei denen aus der Energie- und Winkelanalyse der rückgestreuten Ionen Rückschlüsse über die Identität der Oberflächenatome und ihre Anordnung gewonnen werden können. Die experimentellen Daten lassen sich einfach im Rahmen eines Billardkugelmodells für die Atome interpretieren, wobei Abschattungseffekte sehr empfindlich von der genauen Position der Atome abhängen. Niederenergetische Ionen ($E < 2$ keV), für die jedoch die Neutralisationswahrscheinlichkeit beim Zusammenstoß mit den Festkörperatomen sehr hoch ist, werden nur von Atomen der Oberflächenschicht gestreut und sind daher extrem oberflächenempfindlich.

Feldionenmikroskopie (FIM). Die physikalische Grundlage der von E. W. MÜLLER (1951) entwickelten Feldionenmikroskopie ist die Ionisation neutraler Gasatome in einem elektrischen Feld hoher Feldstärke. Um diese hohe Feldstärke mit praktisch handhabbaren Spannungen von einigen Kilovolt erreichen zu können, muß die Kristallprobe als scharfe Spitze mit einem Radius von 10 bis 50 nm ausgebildet sein. Die Gestalt der Spitze ist dann durch viele Stufen und einzelne hervorragende Atome gekennzeichnet (Abb. 6.10.3–4), die zu einer starken Inhomogenität bzw. Erhöhung des elektrischen Feldes an diesen Stellen führen. Die Ionisation erfolgt bei einer kritischen Distanz x_c oberhalb der Oberfläche (≈ 0,5 bis 1 nm), wobei ein Elektron zur als Anode gepolten Kristallspitze durchtunnelt. Die positiv geladenen Gasionen werden radial auf einen Fluoreszensschirm beschleunigt, wo sie das bis zu zweimillionenfach vergrößerte Bild der Spitzenoberfläche erzeugen. Die höchste praktisch erzielbare Auflösung von 0,2 bis 0,3 nm wird mit He-Atomen bei einem Gasdruck von ≈ 0,1 Pa und tiefen Temperaturen ($T \approx 20$ K) erreicht und ermöglicht so die Abbildung einzelner Atome oder Leerstellen sowie Adsorptions- und Oberflächendiffusionsprozesse.

Die *Atomsonde* ist eine Erweiterung des Feldionenmikroskops (E. W. MÜLLER, 1967). Dabei werden die im FIM-Bild beobachteten Atome durch einen kurzzeitigen Hochspannungsimpuls *(Feldverdampfung)* von der Kristallspitze abge-

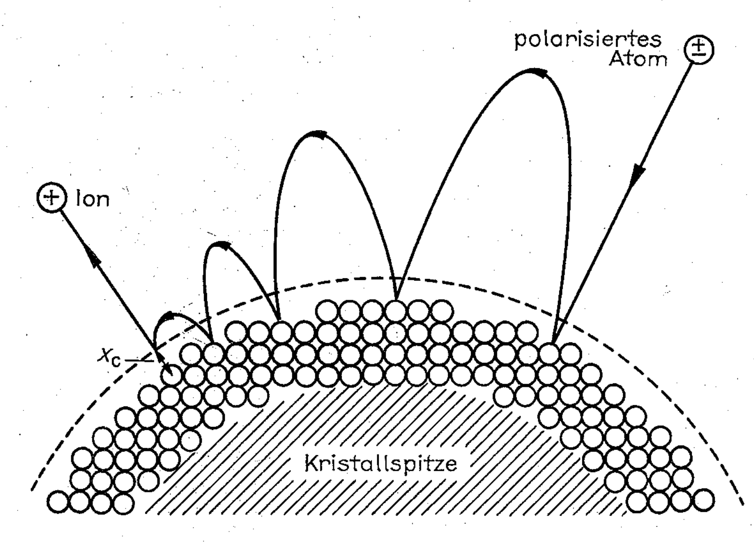


Abb. 6.10.3–4 Zum Abbildungsmechanismus im Feldionenmikroskop

löst und anschließend massenspektroskopisch identifiziert.

Experimentelle Ergebnisse. Die Abstände zwischen den Atomen im Kristallgitter werden neben anderen Faktoren wesentlich von der Zahl der nächsten Nachbarn und der Bindungsart bestimmt. Man erwartet daher für die Oberflächenatome andere Bindungslängen als im Inneren des Kristalls. Die bisher zum größten Teil durch LEED erhaltenen Resultate zur Oberflächenkristallografie lassen für die einzelnen Substanzklassen und Oberflächen ein graduelles Schema an Änderungen erkennen, das bei einer idealen Begrenzung des Kristalles anfängt (keine Differenzen zur Struktur im Volumen) und über erhöhte Atomschwingungen und Relaxationen des Netzebenenabstandes bis zur Umordnung der Atome *(Rekonstruktion)* reicht.

Metalle. In den meisten Fällen nehmen die Metallatome auch an der Oberfläche die Gitterpositionen ein, die sie im Inneren des Kristalls besetzen. Die dichtgepackten Oberflächen behalten die vom Volumen her bekannten Bindungslängen bei. Zu einer Verkürzung der Bindungslänge zwischen den Atomen der obersten beiden Schichten um 2 bis 4% kommt es bei den weniger dicht gepackten kubisch-flächenzentrierten (110) bzw. kubisch raumzentrierten (100) und (111)-Oberflächen. Bei einigen Metalloberflächen (z.B. Ir, Pt, Au, V, Cr, Mo, W) kommt es zu einer Umordnung der Atome der obersten Schicht relativ zu den Gitterpositionen der darunterliegenden Netzebenen. Obwohl eine endgültige Klärung dieser Phänomene noch fehlt, scheinen die Oberflächenatome bestrebt zu sein, möglichst dicht gepackte Strukturen zu bilden.

Halbleiter. Eine viel stärkere Störung als bei Metallen stellt wegen der gerichteten Bindung zwischen den Atomen die Oberfläche von Halbleitern dar. Bei den *Elementhalbleitern* Silicium und Germanium sind die (111)-Oberflächen am stabilsten, die geringe Reichweite der rein kovalenten Bindung führt jedoch zu Veränderungen in den Bindungsorbitalen im Oberflächengebiet. Die Folge sind Umordnungen der Atome bis zu vier Schichten in den Kristall hinein. Bei den binären *Verbindungshalbleitern* übt der weitreichende Anteil der ionischen Bindung einen stabilisierenden Einfluß auf die nichtpolaren Oberflächen aus. Die Dimensionen der Elementarmaschen der bisher untersuchten (110)-Zinkblende-Oberflächen sind die gleichen wie im Volumen, jedoch tritt im Oberflächengebiet eine Scherung zwischen dem Kationen- und Anionengitter auf, wobei die Anionen nach außen und die Kationen nach innen relaxieren. Als Folge kommt es zu umfangreichen Bindungsrotationen und Änderungen der Bindungslängen in den ersten zwei Schichten. Die Atomanordnung der polaren Oberflächen ist viel komplexer und hängt von der Präparationsmethode und der thermischen Behandlung ab.

6.10.4. Elektronenstruktur an der Oberfläche

Die Oberfläche eines kristallinen Festkörpers unterbricht die Periodizität des Kristallpotentials. Außerhalb des Kristalls geht das Potential nach dem Bildkraftgesetz zu Null, und es entsteht an der Oberfläche eine Potentialbarriere *(Oberflächenpotential)*, die die im Kristall frei beweglichen Elektronen nicht heraustreten läßt (Abb. 6.10.4–1). Die Höhe dieser Potentialbarriere ist materialabhängig und läßt sich durch die volumenspezifische *Fermi-Energie* E_F (vgl. 6.3.1.) und die oberflächenspezifische *Austrittsarbeit* Φ angeben. Als zweite wesentliche Folge des Abbruchs der Periodizität des Kristallpotentials können Elektronen an der Oberfläche lokalisiert sein. Dieses Konzept der *Oberflächenzustände* folgt aus quantenmechanischen Betrachtungen und wurde lange vor den experimentellen Nachweisen von TAMM 1932 und SHOCKLEY 1939 begründet. Da elektronische und strukturelle Eigenschaften der Festkörper voneinander abhängen, gehen mit den Rekonstruktionen in der Oberfläche drastische Änderungen der Elektronenstruktur einher. Informationen über die Elektronenstruktur an Oberflächen werden hauptsächlich mit Hilfe von oberflächenempfindlichen Methoden der *Elektronenspektroskopie* erhalten, wobei die kinetische Energie der emittierten oder gestreuten Elektronen analysiert wird. Beispiele dafür sind die *Ultraviolett-Fotoemissionsspektroskopie (UPS), Feldemissionsspektroskopie (FES), Ionenneutralisationsspektroskopie (INS), Augerelektronenspektroskopie (AES)* und die verschiedenen Arten der *Energieverlustspektroskopie (ELS)*. Darüber hinaus ist die Absorption von elektromagnetischer Strahlung (Licht), deren Energie kleiner als die Energiebandlücke bei Halbleitern ist, eine geeignete Methode zur Untersuchung der möglichen elektronischen Übergänge zwischen Oberflächenzuständen untereinander oder zwischen ihnen und dem Valenz- bzw. Leitungsband.

6.10.4.1. Austrittsarbeit. Als Austrittsarbeit bezeichnet man die Minimalenergie, die gebraucht wird, um ein Elektron aus einem Metall herauszuziehen. Im wesentlichen wird der Wert der Austrittsarbeit von den *Austausch-* und *Korrelationseffekten* (vgl. 6.3.1.) der Elektronen untereinander bestimmt. Ein zweiter wichtiger Faktor ist die Dipolschicht an der Oberfläche, die ihre Ursache darin hat, daß die Elektronenwellenfunktionen exponentiell ins Vakuum abklingen (vgl. Abb. 6.10.4–1b, c). Dadurch reicht ein gewisser Teil der Elektronendichte über den eigentlichen geometrischen Ort der Oberfläche hinaus. Dieser Anteil fehlt dann zur Kompensation der positiv geladenen Ionenrümpfe in der Oberflächenschicht. Für die verschiedenen Elemente liegen die Werte von Φ im Bereich von 2 bis 6 eV. Geringe Differenzen $\Delta\Phi < 1$ eV treten für die einzelnen Kristalloberflächen des gleichen Elements auf. Diese *Anisotropie der Austrittsarbeit* wird durch die unterschiedliche Packungsdichte der Oberflächen hervorgerufen. Zahlreiche empirische Untersuchungen haben Zusammenhänge zwischen der Austrittsarbeit und den atomaren oder kristallinen Eigenschaften der Elemente, wie *Atomzahl* und *-volumen, Ionisierungsenergie, Oberflächenenergie* und *-morphologie, Elektronegativität* u. a., aufgedeckt. Die experimentelle Bestimmung der Austrittsarbeit erfolgt entweder direkt durch Analyse thermisch, foto- oder feldemittierter Elektronen oder indirekt durch Messung der *Kontaktpotentialdifferenz*, die bei elektrischer Verbindung zwischen der Probe und einer Referenzelektrode existiert, deren Austrittsarbeit bekannt ist.

6.10.4.2. Oberflächenzustände. Das Hauptmerkmal der *Oberflächenzustände (OF-Zustände)* ist, daß ihre Energie in den verbotenen Zonen der Bandstruktur des unbegrenzten Gitters (vgl. 6.3.8.1.) liegen. Sie spielen daher eine große Rolle für die elektrischen und optischen Eigenschaften von *Halbleitern* und *Isolatoren*. Die Amplituden ihrer Wellenfunktionen haben ihr Maximum in der Oberfläche und fallen exponentiell zu beiden Seiten der Oberfläche hin ab, sie sind daher in der Oberfläche lokalisiert. OF-Zustände, die sich allein auf die Begrenzung des Gitters zurückführen lassen, bezeichnet man als *intrinsic-(Eigen-)Zustände*, werden sie dagegen durch die Existenz von Fremdatomen in der Oberfläche hervorgerufen, spricht man von *extrinsic-(Stör-*

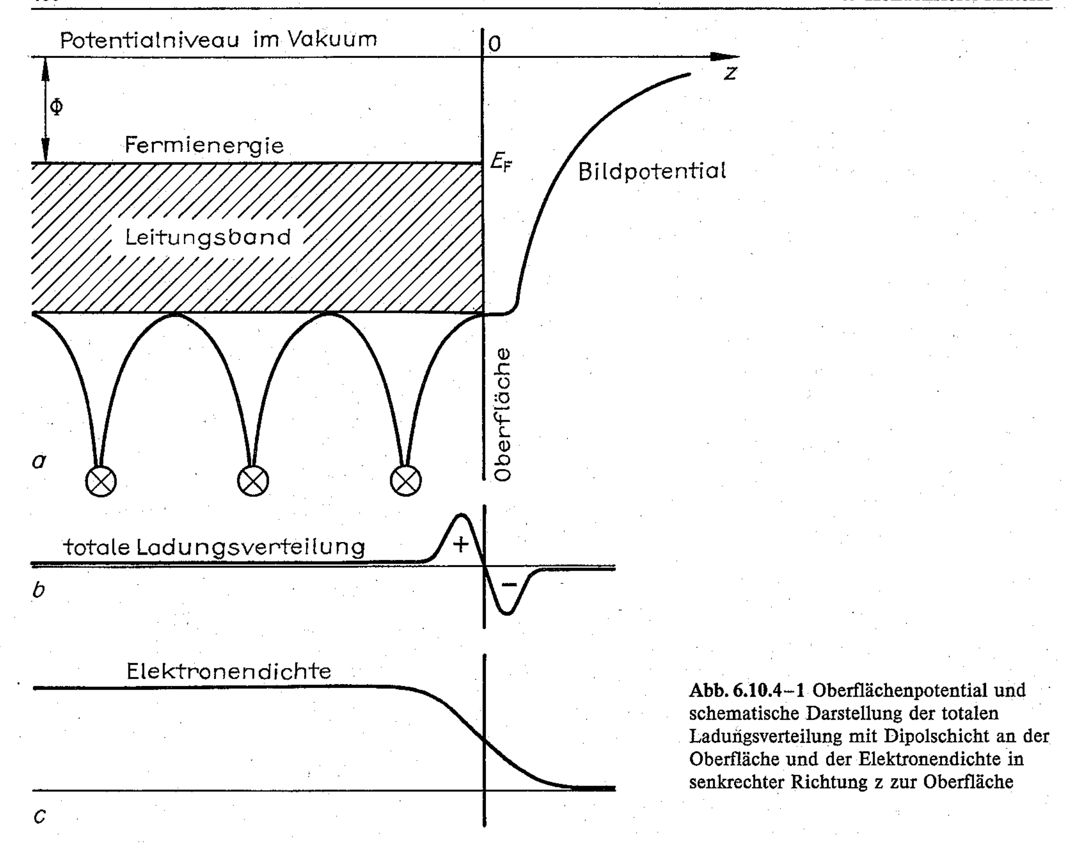


Abb. 6.10.4–1 Oberflächenpotential und schematische Darstellung der totalen Ladungsverteilung mit Dipolschicht an der Oberfläche und der Elektronendichte in senkrechter Richtung z zur Oberfläche

stellen-)Zuständen. Beide Arten der OF-Zustände können als *Donatoren* oder *Akzeptoren* (vgl. 6.3.8.3.) wirken. Die von den OF-Zuständen stammende zusätzliche elektrische Ladung wird aus Gründen der notwendigen Ladungsneutralität durch entgegengesetzte Ladungsträger kompensiert, die sich in einem Gebiet $\gtrsim 100$ nm Dicke unterhalb der Oberfläche des Halbleiters befinden. Solch eine *Raumladungsschicht* stellt sich auch bei Einwirkung eines äußeren elektrischen Feldes oder in unmittelbarer Nähe eines anderen Festkörpers mit unterschiedlicher Austrittsarbeit ein. Als Folge tritt in diesem Gebiet senkrecht zur Oberfläche eine auf- oder abwärts gerichtete Verbiegung der im Volumen horizontal verlaufenden Energiebänder auf. Die experimentelle Bestimmung der *Bandverbiegung* stellt den Zugang zur quantitativen Analyse der OF-Zustände dar.

6.10.5. Physikalische Phasenumwandlungen an Festkörperoberflächen

Phasenumwandlungen an Oberflächen (fest-flüssig, fest-gasförmig) erfordern im Vergleich zu der bei der Phasenentstehung notwendigen homogenen Keimbildung nur geringe Abweichungen vom thermodynamischen Gleichgewicht. Ihre Kinetik wird durch Transportprozesse zur Oberfläche und entlang der Oberfläche und der Ableitung bzw. Zufuhr der Umwandlungswärmen bestimmt. Oberflächen amorpher Festkörper und Kristalloberflächen hoher Indizierung, d. h. hoher Dichte atomarer Stufen, sind atomistisch gesehen permanent rauh und zeigen *kontinuierliches Wachstum*, bei dem ankommende Atome dort eingebaut werden, wo sie auftreffen. Dagegen wird das Wachstum dichtgepackter, niedrig indizierter Kristalloberflächen in analoger Weise wie die Phasenentstehung durch Keimbildungsvorgänge gehemmt.

6.10.5.1. Kristallwachstum aus der Dampfphase. Das Kristallwachstum aus der Dampfphase ist das theoretisch und experimentell am besten untersuchte System und die an ihm von Kossel und Stranski entwickelten molekularkinetischen Vorstellungen haben sich auch bei der Übertragung auf andere Phasengrenzflächen (fest-flüssig, fest-Lösung) bewährt. Bei der Anlagerung von Atomen auf der Oberfläche ergeben sich die in Abb. 6.10.5–1 gezeigten verschiedenen Anlagerungsmöglichkeiten auf der Fläche *1*, an der Stufe *2*, an der sogenannten Halbkristallage *3*, in einer Stufenlücke *4* und in einer Leerstelle *5*, die in dieser Reihenfolge mit wachsenden freiwerdenden Bindungsenergien verbunden sind. Die Halbkristallage *3* spielt die entscheidende Rolle, da bei der Anlagerung (Ablösung) eines Atomes an (aus) ihr die Bindungsenergie des Kristalles je Atom frei (gebraucht) wird und dieselbe geometrische Konfiguration erhalten bleibt, wodurch dieser „wiederholbare Schritt" am häufigsten realisiert wird. Das *Wachstum an perfekten*, stufenfreien *Kristalloberflächen* verläuft über die folgenden Elementarschritte: 1. Aus der Dampfphase einfallende Atome werden thermisch akkomodiert und führen als

adsorbierte Atome (Adatome) häufige Platzwechsel (Oberflächendiffusion) aus, bevor sie entweder wieder verdampfen oder sich mit anderen diffundierenden Adatomen zusammenlagern. 2. Die durch statistische Schwankungen entstehenden zweidimensionalen Flächenkeime besitzen unterhalb einer bestimmten kritischen Größe größere Zerfalls- als Wachstumswahrscheinlichkeiten. Solch ein Keimbildungsvorgang ist nur möglich bei einer bestimmten Abweichung vom thermodynamischen Gleichgewicht. Im Falle des Wachstums aus der Dampfphase ist die Übersättigung $(p - p_0)/p_0 > 1$ für die Größe des kritischen Keimes, seine Bildungsrate und damit auch für die Wachstumsgeschwindigkeit der Kristalloberfläche bestimmend (p_0 = Sättigungsdampfdruck, p = aktueller Dampfdruck). 3. An die Randstufen stabiler Keime lagern sich diffundierende Adatome an und gelangen durch Randdiffusion entlang der Stufen zu den effektiven Wachstumsstellen (Halbkristallagen). 4. Die Stufen wachsen zu den Flächenrändern aus oder annihilieren sich mit anderen Stufen, worauf erneute Keimbildung einsetzt.

Für das *Wachstum an realen Kristalloberflächen*, die als Kristallbaufehler, z. B. Schrauben- und Stufenversetzungen enthalten (vgl. Abb. 6.10.5–1), werden wesentlich geringere Übersättigungen benötigt. Eine Schraubenversetzung wirkt als ein kontinuierlicher Flächenkeim und die mit ihr verbundene Stufe windet sich zu einer sogenannten Wachstumsspirale auf. In bestimmten Fällen sind auch Stufenversetzungen geeignet, die für die Bildung von Flächenkeimen notwendige Keimbildungsarbeit herabsetzen.

6.10.5.2. Verdampfung von Kristalloberflächen. Die *Verdampfung* verläuft nach den gleichen Elementarschritten wie das Wachstum, nur mit umgekehrtem Vorzeichen für den Strom der Atome. Auf realen Oberflächen entstehen an Schraubenversetzungen Abdampfspiralen und an Stufenversetzungen konzentrische Kreisstufen, und als weitere kontinuierliche Stufenquellen kommen noch die Flächenränder hinzu.

6.10.5.3. Wachstum dünner Schichten aus der Dampfphase. Das Wachstum dünner Schichten unterscheidet sich von dem in 6.10.5.1. behandelten homogenen Fall in zweierlei Hinsicht: 1. Es liegt der heterogene Fall der Abscheidung auf artfremder Unterlage vor, bei dem die Bindungsenergien der Depositatome untereinander und zur Unterlage sehr unterschiedlich sein können. 2. Das Schichtwachstum wird praktisch nur bei sehr hohen Übersättigungen durchgeführt, wodurch der kritische Keim einem Einzelatom bzw. der Zusammenlagerung weniger Atome [engl. cluster] entspricht. Für das Zusammenwachsen (Koaleszenz) der gebildeten Cluster spielt im Anfangsstadium ihre Beweglichkeit (Oberflächendiffusion) eine wichtige Rolle, während bei der späteren Bildung geschlossener Schichten die Berührungskoaleszenz überwiegt, bei der eine spontane Rekristallisation zwischen verschieden orientierten Inseln (sogenannte flüssigkeitsähnliche Koaleszenz) auftreten kann. In Abhängigkeit von der Bindungsenergie innerhalb des Deposits bzw. zwischen Deposit und Unterlage kann man qualitativ drei Grenzfälle für das Wachstum geschlossener Schichten unterscheiden: 1. dreidimensionales oder *Inselwachstum*, wenn die Bindung innerhalb des Deposits stärker als zur Unterlage ist, 2. zweidimensionales oder *Monoschichtwachstum*, wenn die Bindung zur Unterlage stärker als die innerhalb des Deposits ist, und 3. den *Stranski-Krastanov-Mechanismus*, bei dem anfängliches Monoschicht- in Inselwachstum übergeht.

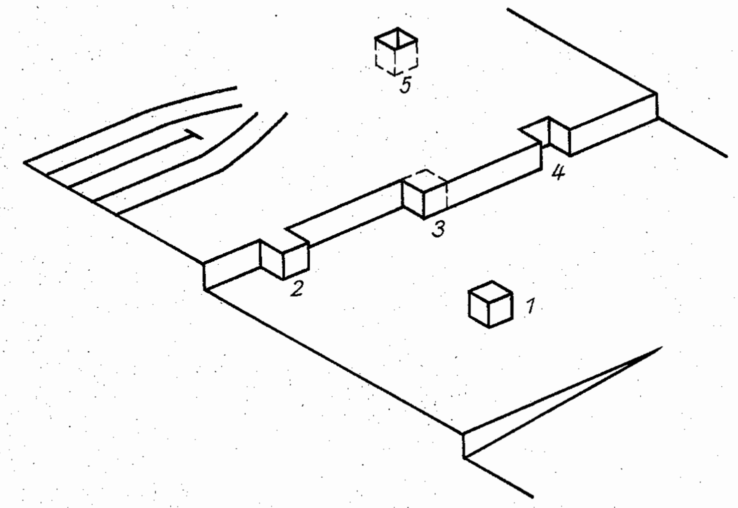


Abb. 6.10.5–1 Unterschiedliche Bindungsplätze an einer Kristalloberfläche und Durchstoßpunkte einer Stufenversetzung *(links)* bzw. einer Schraubenversetzung *(rechts)*

6.10.6. Adsorption auf Festkörperoberflächen

6.10.6.1. Grundlagen. Zahlreiche technologische Prozesse wie auch chemische Reaktionen setzen die Anlagerung von Atomen oder Molekülen *(Adsorbate)* auf Festkörperoberflächen *(Substrate)* voraus. Dieser Vorgang *(Adsorption)* läuft exotherm ab und stellt die allgemeinste Form der Reduzierung der freien Oberflächenenergie dar. Nähert sich ein Atom oder Molekül aus der Gasphase der Festkörperoberfläche, so erfährt es aufgrund elektrischer Kräfte ein anziehendes Potential, bis Elektron-Elektron-Wechselwirkungen oder schließlich die Kernabstoßung eine weitere Annäherung verhindern (Abb. 6.10.6–1a). Der *Adsorptionsplatz* wird durch die Potentialmulde bestimmt. Ihre Tiefe gibt die freiwerdende *Adsorptionswärme* ΔH_{ad} an und ist ein Maß für die Stärke der Wechselwirkung. Bei der *Physisorption* ($\Delta H_{ad} < 0{,}5$ eV) beruht diese lediglich auf van-der-Waals-Kräften ($V \sim r^{-6}$) zwischen den Oberflächenatomen und dem Adsorbat, wobei die Anziehung hauptsächlich durch induzierte fluktuierende Ladungsverteilungen *(Dispersions-*

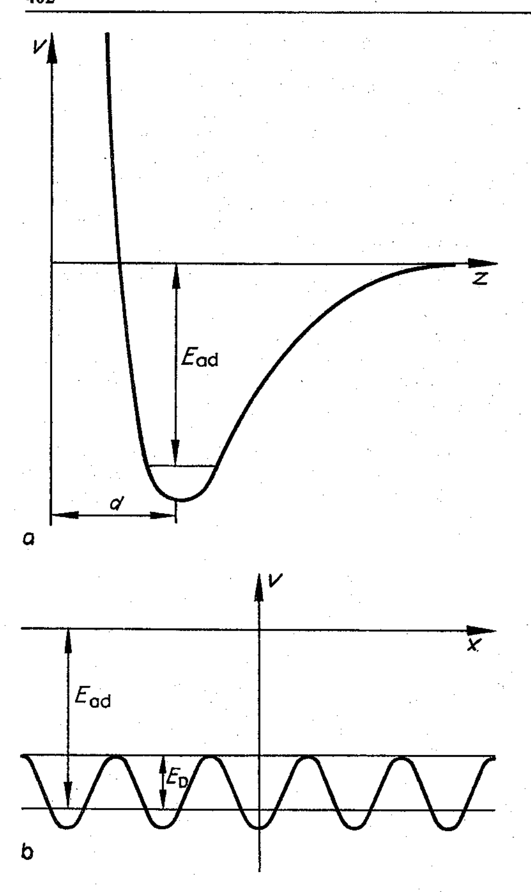


Abb. 6.10.6–1 Potentialverlauf für Adsorption eines Atomes oder Moleküles mit der Adsorptionsenergie E_{ad} auf einer kritstallinen Festkörperoberfläche; E_D ist die Aktivierungsenergie für Oberflächendiffusion, (a) senkrecht und (b) im Abstand $z = d$ parallel zur Oberfläche

kräfte) hervorgerufen wird. Erfolgt dagegen in einem gewissen Umfang eine Umverteilung der Elektronen zwischen den an der Adsorption beteiligten Atomen in Richtung einer kovalenten oder ionischen Bindung, so führen die starken elektrostatischen Kräfte ($V \sim r^{-1}$) zu $\Delta H_{ad} > 0{,}9$ eV. Wegen des chemischen Charakters der Bindung spricht man hier von *Chemisorption*, die *Bindungsenergien* betragen in den meisten Fällen 1 bis 5 eV. Beide Arten der Adsorption treten bei der *Dissoziation* mehratomiger Moleküle auf, wo das Molekül zuerst als Ganzes physisorbiert wird, während bei weiterer Annäherung an die Oberfläche die Molekülbindung aufgebrochen wird und individuelle Atome oder Radikale einzeln im chemisorbierten Zustand gebunden werden. Die adsorbierten Partikeln werden anfänglich mit der *Adsorptionsenergie* E_{ad} gebunden. Mit der Zeit nehmen sie durch Schwingungen jedoch so viel Energie vom Gitter auf, daß sie die Potentialbarriere wieder überwinden können. Ihre Verweilzeit auf der Oberfläche wird durch das Verhältnis von Adsorptionsenergie und thermischer Energie $k_B T$ (vgl. 6.4.) bestimmt, $\tau = \tau_0 \exp[E_{ad}/(k_B T)]$ ($\tau_0 \sim 10^{-13}$ s = Zeitdauer für eine einzelne Atomschwingung). Eine ausreichende Experimentierzeit wird bei der Physisorption wegen der geringen Adsorptionsenergie nur bei tiefen Temperaturen erreicht. Die hohen Bindungsenergien der Chemisorption bewirken praktisch eine unendliche Verweilzeit des Adsorbates. Parallel zur Kristalloberfläche wird das Potential periodisch entsprechend der Atomanordnung moduliert (Abb. 6.10.6–1b). Die sich wiederholenden Potentialbarrieren geben die Aktivierungsenergien E_D des Adsorbates für die *Oberflächendiffusion* (Platzwechselvorgänge) an. Für einzelne Atome ist $E_D \leqq E_{ad}/4$, so daß sie ihren Adsorptionsplatz viele Male wechseln, ehe sie desorbieren. Die Adsorbate bewegen sich relativ frei über die Oberfläche, wenn ihre thermische Energie die Diffusionsenergie übersteigt ($k_B T > E_D$). In diesem Fall liegt eine nichtlokale Adsorption vor.

6.10.6.2. Geometrische Struktur und Adsorptionsplätze chemisorbierter Phasen. Mit wachsender Bedeckung spielt neben der Adsorbat-Substrat-Wechselwirkung auch diejenige zwischen den adsorbierten Partikeln eine entscheidende Rolle. Die Bindungs- und Diffusionsenergie, die Adsorptions- und Desorptionskinetik, die Ladungsverteilung bei der Chemisorption, vor allem aber die relative Anordnung der Adsorbatatome untereinander und relativ zu den Substratatomen, hängen empfindlich vom *Bedeckungsgrad* Φ ab. Dipol-Dipol-Kräfte zwischen den adsorbierten Partikeln, Überlappung der Elektronenorbitale bei enger Nachbarschaft sowie eine indirekte Wechselwirkung über die Leitungselektronen der Substratunterlage können das Potentialfeld der perfekten Einkristalloberfläche so beeinflussen, daß sich anstatt einer regellosen Anordnung der adsorbierten Partikeln geordnete zweidimensionale Phasen mit äquidistanten Atomabständen herausbilden. Diese Anordnung steht dann häufig in enger Beziehung zur Periodizität des darunterliegenden Substratgitters. So besetzen in vielen Fällen bei Bedeckungen $\Phi \leqq 0{,}5$ die Adsorbatatome entlang bestimmter kristallografischer Richtungen nur jeden zweiten Gitterplatz der Oberfläche (Abb. 6.10.6–2). In einigen Substrat/Adsorbat-Systemen fallen der *Haftkoeffizient* und die Adsorptionswärme bei $\Phi = 0{,}5$ so stark ab, daß eine weitere Adsorption nicht mehr erfolgen kann. Überwiegt die Wechselwirkung innerhalb der Adsorbatschicht bei weitem diejenige zum Substrat, wie es bei der Physisorption von Edelgasen zutrifft, so beeinflußt das Oberflächengitter kaum die geometrische Struktur der Adsorptionsschicht.

Adsorptionsplätze. Auf einer Kristalloberfläche existieren mehrere, nicht gleichwertige Positionen, die zu Potentialmulden für ein auftreffendes Atom oder Molekül nach Abb. 6.10.6–1a gehören, sich aber in der Zahl der nächsten Nachbarn *(Koordinationszahl)* unterscheiden (vgl. Abb. 6.10.6–2). Welche Positionen tatsächlich als Adsorptionsplätze fungieren, hängt auch von der Wechselwirkung zwischen den Adsorbatatomen und der Natur der chemischen Bindung ab. Die bisher vorwiegend mit Elektronenbeugung (LEED, vgl. 6.10.3.3.) untersuchten Chemisorptionssysteme liefern unter Berücksichtigung weniger Ausnahmen folgende allgemeine Regeln:

1. Adsorptionsplatz wird derjenige, der den Adsorbatatomen die größtmögliche Zahl nächster Substratatome anbietet. Daraus folgt, daß die Adsorptionsplätze zumeist auch Plätze höchster Symmetrie sind und zu den normalen Gitterplätzen des Substrats gehören.
2. Der Abstand zwischen Metall- und Adsorbatatom unterscheidet sich nur geringfügig von dem des gleichen Paares im Molekül- oder Festkörperverband.

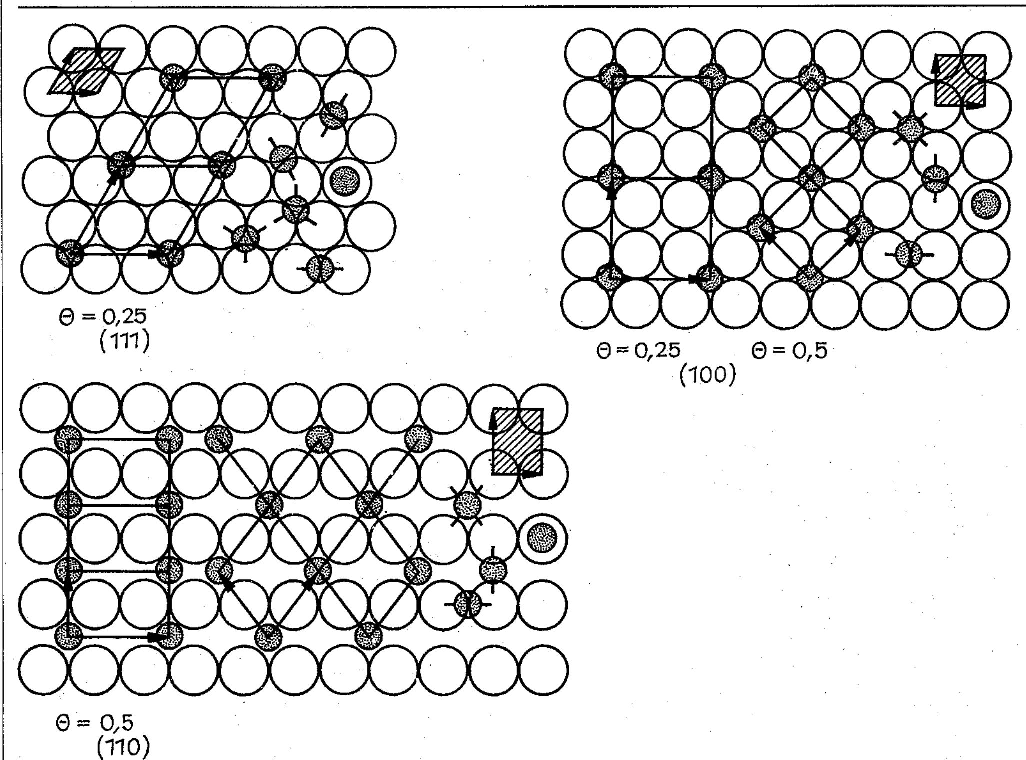


Abb. 6.10.6–2 Atomanordnung (schwarze Punkte) geordneter adsorbierter Phasen des Bedeckungsgrades *Θ* und Adsorptionsplätze unterschiedlicher Koordinationszahl auf (111)-, (100)- und (110)-Oberflächen des kubisch-flächenzentrierten Gitters. Die Elementarmasche des Substratgitters ist schraffiert gezeichnet

3. Adsorbate können Rekonstruktionen der Substratoberfläche (vgl. 6.10.3.3.) aufheben oder hervorrufen.

6.10.6.3. Elektronenstruktur von Adsorbatsystemen. Der Umfang des Elektronentransfers zwischen Adsorbat- und Substratatomen bei der Chemisorption hängt von der *Elektronenaffinität* (freiwerdende Energie bei der Anlagerung eines Elektrons an ein isoliertes Atom) und der *Ionisierungsenergie* des Adsorbates ab. Ist die Austrittsarbeit des Substrats kleiner als die Elektronenaffinität bzw. größer als die Ionisierungsenergie, so erfolgt ein Ladungstransfer im ersten Fall in Richtung des Adsorbats, im zweiten Fall in Richtung des Substrats. Als Folge stellt sich eine Dipolschicht an der Oberfläche ein, die eine Änderung der Austrittsarbeit der unbedeckten Substratoberfläche bewirkt. Diese Änderung ist eine Funktion der Bedeckung. Der Ladungstransfer pro adsorbiertes Atom übersteigt bei der Chemisorption jedoch nicht 10% der Elementarladung eines Elektrons. In Extremfällen kann aber unter den erwähnten Bedingungen eine positive oder negative Ionisation des adsorbierten Atoms erfolgen. Dieses Atom wird dann als Ion von der Oberfläche wieder emittiert *(Oberflächenionisation).*

7. Meßmethoden der Physik

7.1. Labortechnik

Seit dem Entstehen von Laboratorien für Experimentalphysik hat sich aus einfachen Anfängen eine Labortechnik entwickelt, die heute recht umfangreiche Spezialkenntnisse verlangt. Der Begriff Labortechnik ist allerdings nicht genau fixiert und wird im weiteren Sinne für die Anwendung der Technik bei Laborversuchen, also von Verfahren, Technologien und Geräten, gebraucht. Ein neu erdachtes Experiment erfordert meist neue technische Mittel oder Änderungen an bestehenden Apparaturen, wenn es sich nicht um routinemäßige Anwendung eines Verfahrens auf andere Substanzen handelt. Diese Hilfsmittel werden von Physikern u.a. Wissenschaftlern oft selbst konstruiert, unter Umständen berechnet, und in Zusammenarbeit mit Werkstätten, Ingenieuren und Laboranten entwickelt und erprobt. Erst wenn sich aus solchen Arbeiten, die in vielen Laboratorien aufgegriffen werden, gewisse Routineverfahren herausgebildet haben und die zugehörigen apparativen Hilfsmittel industriell hergestellt werden, wird man von einer Labortechnik im engeren Sinne sprechen, z. B. der Vakuum- oder Tieftemperaturtechnik. Auch der Begriff Laboratorium wird unterschiedlich angewendet, die Größe eines Labors kann von einem mehr oder weniger zweckmäßig ausgestatteten Zimmer bis zur riesigen Halle, dem Turm oder dem Tunnel eines Beschleunigers für Kern- oder Elementarteilchenphysik mit angrenzenden Gebäuden und Meßplätzen reichen (verschiedentlich wird unter Laboratorium ein größeres Institut mit weitgespannter Forschungsthematik verstanden, z.B. beim bekannten Cavendish Laboratory in Cambridge, England).

7.1.1. Erzeugung tiefer Temperaturen

7.1.1.1. Laboranwendungen tiefer Temperaturen. Seit der Gründung spezieller Tieftemperaturlaboratorien (Kammerlingh-Onnes in Leyden u.a.) hat sich die Anwendung tiefer Temperaturen immer mehr durchgesetzt, so daß heute physikalische Laboratorien ohne eine entsprechende Technik kaum noch vorstellbar sind. Unter tiefen Temperaturen im weiteren Sinne werden Temperaturen unterhalb des Gefrierpunktes von Wasser verstanden, hauptsächlich denkt man jedoch an den Bereich der Siedepunkte von Gasen, wie Stickstoff, Sauerstoff, Wasserstoff, den Edelgasen, und darunter.

Tiefe Temperaturen werden in großem Maße für Meßzwecke benutzt, wobei die Temperaturabhängigkeit zahlloser Eigenschaften aus theoretischen Gründen (Vergleich mit physikalischen Modellen) und für praktische Anwendungen (Vakuumtechnik, Supraleitung, Weltraumtechnik) von Interesse ist. Im Gebiet extrem tiefer Temperaturen werden immer noch neuartige physikalische Phänomene beobachtet, z. B. Ordnungszustände und Phasenumwandlungen gewisser Eigenschaften von Atomkernen (Kernspins). Da der Wasserdampf der Luft sofort an den kalten Oberflächen kondensiert und durch Neben- oder Kurzschlüsse sowie Beschlagen von Fenstern z. B. elektrische oder optische Messungen stören würde, werden die Experimente fast stets im Vakuum unter Benutzung von oft komplizierten Kryostaten ausgeführt. Tiefe Temperaturen selbst werden in großem Umfang bei der Vakuumerzeugung eingesetzt (vgl. 7.1.3.), um Treibmitteldämpfe und Verunreinigungen auszufrieren oder eine Pumpwirkung durch Adsorption zu erzielen. Verwandt damit ist die Benutzung zur Stofftrennung und Reinigung, z. B. die Herstellung von Sauerstoff und Stickstoff sowie die Edelgasgewinnung aus verflüssigter Luft, die großtechnisch mittels fraktionierter Destillation erfolgt. Die Weltraumsimulation stellt ein Anwendungsbeispiel der Materialprüfung bei tiefen Temperaturen dar. Eine wichtige Laboranwendung ist heute auch die Erzeugung des supraleitenden Zustandes entsprechend konstruierter Magnete.

7.1.1.2. Erzeugung tiefer Temperaturen im Bereich der Kältetechnik. Für einfache Anwendungen werden sogenannte Kältemischungen eingesetzt, die als eutektische Gemische eine relativ konstante Temperatur ergeben, z. B. $-21{,}2\,°C$ für 100 g Eis mit 31 g $NaCl \cdot 2H_2O$ oder $-55\,°C$ mit 143 g $CaCl_2 \cdot 6H_2O$. Günstig ist die Anwendung von *Trockeneis* (feste „Kohlensäure", CO_2), das zerstoßen mit einem Überschuß von Alkohol eine Kältemischung für $-78{,}5\,°C$ ergibt. Als Badflüssigkeit bis herab zu $-150\,°C$ läßt sich Pentan verwenden. Werden solche Temperaturen

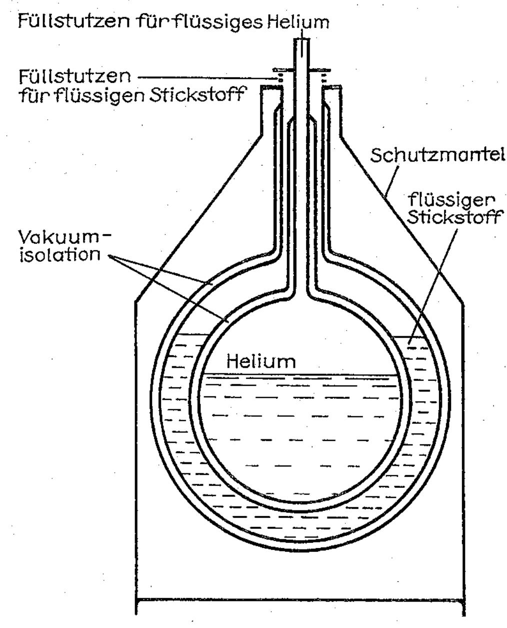


Abb. 7.1.1–1 Doppeldewargefäß für flüssiges Helium oder flüssigen Wasserstoff

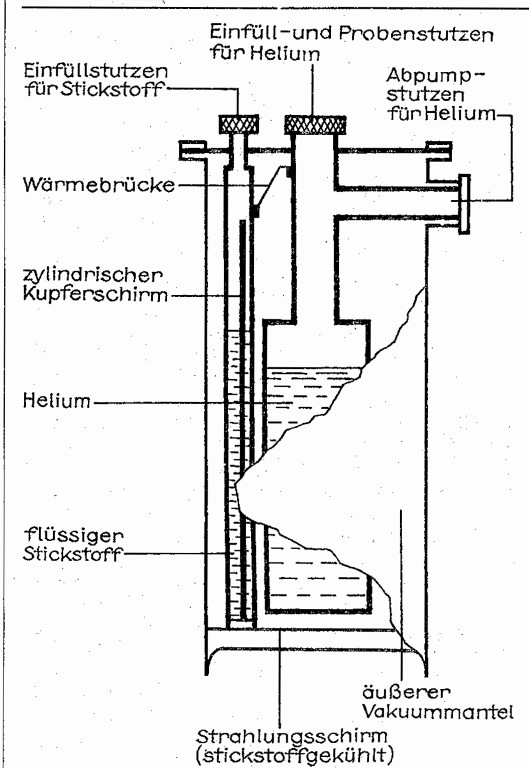
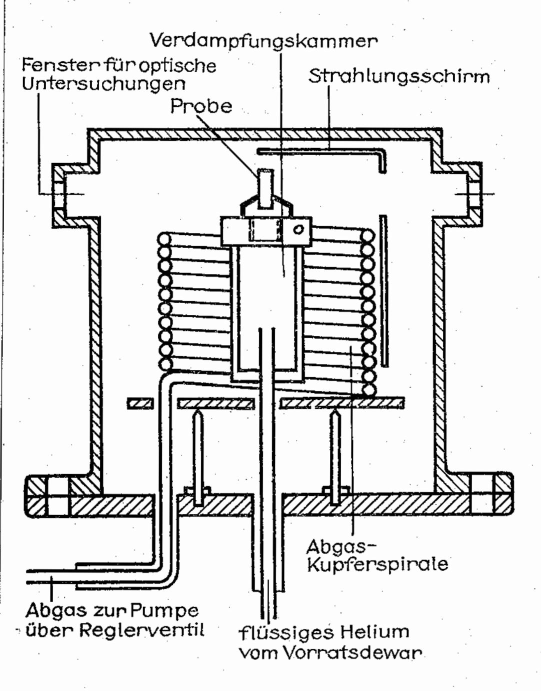


Abb. 7.1.1–2 Heliumbadkryostat (Vakuumpumpstutzen mit Ventil, Meßfühler u. a. weggelassen)



ständig bei einem größeren Wärmeanfall gebraucht, sind Kältemaschinen vorzuziehen, die nach dem Prinzip des Kompressorkühlschrankes arbeiten, wobei als Kältemittel häufig *Freon* benutzt wird. Wegen des geringeren Wartungsaufwandes können Peltier-Elemente günstiger sein, die unter Ausnutzung des Peltier-Effektes mit oder ohne zusätzliche Wasserkühlung eine Temperaturdifferenz bei Gleichstromdurchgang durch eine Halbleiteranordnung erzeugen. Peltier-Batterien werden z. B. für die Kühlung von Leistungsbauelementen, Mikroskoptischen und als Öldampfsperre in Öldiffusionspumpen eingesetzt. Für Meßzwecke werden Verdampfungskryostate (vgl. 7.1.1.3.) in diesem Temperaturbereich benutzt. Es sind auch Anordnungen bekannt, bei denen die „Wärmeübertragung" von einem Flüssigkeitsbad (Stickstoff, Helium) über ein Gas (Wasserstoff) erfolgt, dessen Druck durch Abpumpen eingeregelt wird, wodurch sich eine hohe Konstanz der Temperatur des Meßobjektes erreichen läßt.

7.1.1.3. Erzeugung tiefer Temperaturen im Bereich verflüssigter Gase. Für die Vakuumerzeugung (vgl. 7.1.3.) werden *Kühlfallen (Tiefkühlbaffles)* und *Dewar-Gefäße* (Abb. 7.1.1–1) benutzt, in die von Hand oder automatisch flüssiger Stickstoff gefüllt wird. Aus Sicherheitsgründen ist flüssige Luft zu vermeiden, bei der durch vorherrschende Stickstoffverdampfung eine sehr gefährliche Sauerstoffanreicherung erfolgen kann, da die Siedetemperatur des Stickstoffs 77,3 K, die des Sauerstoffs jedoch 90,2 K beträgt. Da der flüssige Stickstoff bei der Luftverflüssigung (zur Sauerstoffgewinnung) in großen Mengen anfällt, wird er meist vom industriellen Erzeuger bezogen, doch werden auch in manchen Laboratorien kleinere Luftverflüssigungsanlagen (Philips-Verflüssiger) betrieben. Flüssiggase müssen zur Vermeidung von zu großen Verdampfungsverlusten in doppelwandigen, vakuumisolierten Vorratsbehältern (Dewar-Gefäße) aufbewahrt werden, die heute meist aus Edelstahl, sonst noch aus Glas hergestellt sind. Flüssiges Helium muß wegen der sehr geringen Verdampfungswärme sogar in Doppel-Dewar-Gefäßen mit einem Mantelgefäß für flüssigen Stickstoff transportiert und bevorratet werden. Durch Abpumpen des verdampfenden Gases über dem Flüssigkeitsspiegel, d. h. beim Sieden unter vermindertem Druck, läßt sich die normale Siedetemperatur noch etwas unterschreiten. Um tiefe Temperaturen für Meßzwecke zu erzeugen, werden sogenannte *Kryostate* verwendet. Sie müssen durch guten Wärmekontakt eine schnelle und effektive Temperaturübertragung auf das Meßobjekt und natürlich den eigentlichen Meßprozeß für große Parameterbereiche ermöglichen, was oft komplizierte Konstruktionen erfordert.

Badkryostate (Abb. 7.1.1–2) haben ein großes Kältemittelreservoir, z. B. für 5 bis 15 l flüssigen Stickstoff und 2 bis 5 l flüssiges Helium, das nach einmaliger (Abkühlung und) Füllung mehrere Stunden (bis zu 2 Tagen) die erforderliche Probentemperatur erzeugt. Durch Abpumpen von Helium können zunächst Temperaturen bis 2,2 K realisiert werden, wobei das ^{4}He supraflüssig wird. Wenn man das Abwandern des Flüssigkeitsfilmes geometrisch behindert, können Temperaturen bis etwa 0,9 K erreicht werden.

Verdampfungskryostate werden laufend mit dem Kältemittel beschickt, indem ständig das entsprechende Gas abge-

Abb. 7.1.1–3 Heliumverdampferkryostat zum Aufsetzen auf Heliumvorratsgefäß (Meßfühler, z. B. Dampfdruckthermometer, im Kammerkopf weggelassen. Abgasgekühlter Schirm nur rechts gezeichnet)

pumpt und dabei auch das Kältemittel angesaugt wird. Durch geeignete Regelung von Zufluß und Abpumpen lassen sich leicht Temperaturen unterhalb und oberhalb des Kühlmittelsiedepunktes einstellen, z. B. zwischen 2,3 und 300 K. Im Gegensatz zu Badkryostaten wird beim Abbruch eines Experimentes kein weiteres Kältemittel verbraucht. Heliumverdampferkryostate (Abb. 7.1.1–3) kommen ohne flüssigen Stickstoff aus, da durch geeignete Konstruktion das Abgas einen Strahlungsschutzschirm kühlt. In *Heliumkombinationskryostaten*, die mit Bad- oder Verdampfungsbetrieb arbeiten können, wird dagegen wie bei Badkryostaten ein äußeres Dewar-Gefäß mit flüssigem Stickstoff beschickt, so daß das Heliumgefäß und seine Zuleitungen gegen die Wärmestrahlung aus der Umgebung geschützt wird, die einen erheblichen Heliumverbrauch hervorrufen würde. *Verbundkryostate* sind Systeme, die als Heliumbadkryostate für eine supraleitende Magnetspule dienen und in deren zylindrischer Öffnung einen Kryostaten für eine unabhängige Probenkühlung enthalten, so daß die Temperaturabhängigkeit von Festkörpereigenschaften in starken Magnetfeldern untersucht werden kann.

^{3}He-Kryostate ermöglichen die Erzeugung von Temperaturen im Bereich von 1,2 bis 0,3 K, wozu das sehr seltene Isotop ^{3}He erforderlich ist, dessen natürliche Häufigkeit nur 10^{-4} % beträgt. Das ^{3}He-Gas wird an einem ^{4}He-Bad (1,5 K) vorgekühlt und dann dosiert abgepumpt. Anstelle dieses Badkryostatenprinzips kann auch ein Joule-Thomson-Zyklus mit ^{3}He durchgeführt werden.

7.1.1.4. Erzeugung extrem tiefer Temperaturen. Um noch tiefere Temperaturen realisieren zu können wie sie vor allem für Untersuchungen in der Festkörperphysik gebraucht werden, müssen andere physikalische Effekte zur Temperaturerniedrigung herangezogen werden. Neben der seit längerem bekannten adiabatischen Entmagnetisierung paramagnetischer Salze, die in einem allerdings einmaligen, nichtstationären Prozeß 2 mK ergibt, sind seit einigen Jahren auch Mischungskryostate entwickelt worden. ^{3}He/^{4}He-Mischkryostate lassen Temperaturen unterhalb 300 mK bis zu ≈ 10 mK durch die ^{3}He-^{4}He-Lösungs- oder Verdünnungskühlung erreichen. Dabei wird in einem mit flüssigem Stickstoff und ^{4}He vorgekühlten Kryostaten unter Anwendung einer komplizierten Technik die „Verdampfungswärme“ von ^{3}He-Atomen bei dem Übertritt aus einer ^{3}He-reichen flüssigen Phase in die sich darüber schichtende ^{4}He-Phase zur Kälteerzeugung verwendet. Es werden Kälteleistungen zwischen einigen Mikrowatt und etwa 100 μW erzielt, wobei ständig ^{3}He abzupumpen ist. Mit der Pomeranchuk-Methode lassen sich instationär 2 mK durch adiabatische Kompression eines Gemisches von festem und flüssigem ^{3}He erreichen. Noch tiefere Temperaturen erhält man durch die adiabatische Kernentmagnetisierung. Die erzielten (Kernspin-)-„Temperaturen“ im Mikrokelvinbereich und darunter entsprechen dann nur noch der Ordnung der Kernspins, während das Gitter, also der feste Körper, eine höhere Temperatur aufweist.

7.1.2. Erzeugung hoher Drücke

7.1.2.1. Laboranwendungen hoher Drücke. Unter hohen Drücken versteht man i. allg. solche ab ≈ 10 MPa (100 at). Im Laboratorium werden erheblich über den Normalluftdruck hinausgehende Drücke recht unterschiedlich eingesetzt. Umfangreiche Anwendungen ergeben sich bei der Bestimmung mechanischer Eigenschaften von Gasen, Flüssigkeiten und festen Körpern, insbesondere der Kompressibilität bei Anwendung eines allseitigen Druckes sowie der Längenänderung fester Körper und Kristalle unter Einwirkung axialer, d. h. in einer Richtung wirkender Druck- und Zugbeanspruchung. Die Skala solcher Untersuchungen erstreckt sich von Problemen der Materialprüfung, wie Druck- und Zugfestigkeit, über die Bestimmung mechanischer bzw. elastischer Konstanten bis hin zu subtilen Einflüssen auf die elektronische Bandstruktur, die sich in Änderungen optischer Eigenschaften fester Körper bemerkbar machen.

Werden bei Druck- oder Zugbeanspruchung die Elastizitätsgrenzen überschritten, erreicht man das technisch wichtige Gebiet des plastischen Verhaltens, das man auf die verschiedenartigen Defekte des Kristallaufbaues zurückführt. Auch das experimentell und theoretisch stark bearbeitete Gebiet der Phasenumwandlungen und des Verhaltens in der Nähe der kritischen Punkte bedarf der Hochdrucktechnik. In der Labortechnik werden hohe Drücke jedoch auch als einfache Hilfsmittel angewendet, um Gase zu komprimieren, z. B. bei der Wiedergewinnung des wertvollen, in der Kryotechnik eingesetzten Heliums, oder um Bleche oder Netze als Elektroden zur Ablenkung oder Analyse geladener Teilchen geeignet zu formen.

7.1.2.2. Erzeugung und Aufbewahrung komprimierter Gase. Für eine sehr geringe Druckerhöhung lassen sich manche mechanischen Vakuumpumpen einsetzen, z. B. Drehschieberpumpen, indem das Gas auf der Vakuumseite angesaugt und an der Gasaustrittsseite entnommen wird. Für die Abfüllung von Gasen in die bekannten Stahlflaschen werden Kompressoren benutzt, die ebenfalls nach dem Prinzip arbeiten, eine angesaugte Gasmenge abzusperren, durch eine Kolbenwirkung zu komprimieren und dann durch Öffnen eines Ventils in die Stahlflasche zu drücken. Helium u. a. Edelgase dürfen nur in grau gestrichene Stahlflaschen abgefüllt werden, während Wasserstoff in roten, Stickstoff in grünen und Sauerstoff in nicht farbig gekennzeichneten Flaschen aufbewahrt wird, auch sind die unterschiedlichen (Links- und Rechts-)Gewinde für den Anschluß der ebenfalls gastypischen Druckminderventile zu berücksichtigen. Je nach den Anforderungen wird man Gase aus Stahlflaschen direkt oder nur nach entsprechenden Reinigungsprozeduren verwenden können, dazu gehö-

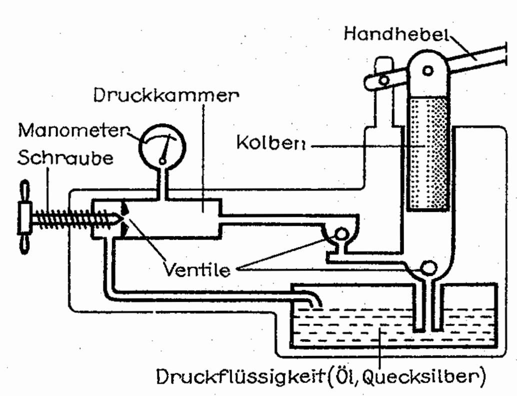


Abb. 7.1.2–1 Cailletet-Pumpe (elektrische Durchführungen bzw. Fenster weggelassen). Beim Heben des Kolbens erfolgt Ansaugen, beim Senken Einpressen der Druckflüssigkeit in die Druckkammer

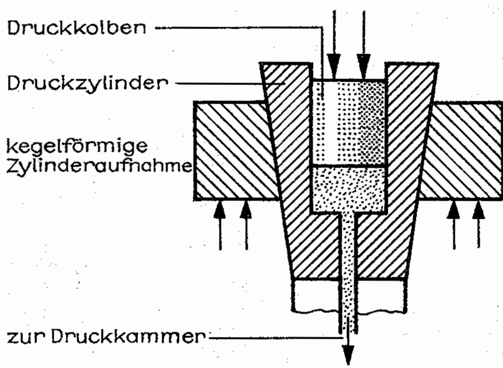


Abb. 7.1.2–2 Prinzip der Höchstdruckpresse nach BRIDGMAN. Beim Einpressen des Druckkolbens wird zugleich der Druckzylinder unter einen wachsenden Außendruck gesetzt, wodurch ein Aufweiten des Kolbens verhindert wird

ren das Ausfrieren von Feuchtigkeitsspuren in Tiefkühlfallen, das Gettern von Gasverunreinigungen durch geeignete, frisch aufgedampfte Metallschichten oder die Diffusion durch Palladiumröhrchen (Wasserstoff) oder Quarzröhren, die auf eine hohe Temperatur erhitzt werden müssen. Für sehr reine Gase lassen sich auch Quecksilberpumpen zur Kompression verwenden, die periodisch durch Druckluft gesteuert werden.

7.1.2.3. Erzeugung hoher und sehr hoher Drücke. Gasdrücke bis 100 MPa lassen sich mit mehrstufigen Kolbenkompressoren erzeugen, durch Verwendung von Druckübersetzern ist noch eine Steigerung auf den zehnfachen Wert möglich. Mit mechanischen (Hebel-)-Pressen oder durch Anwendung des hydraulischen Prinzips erzielt man stationäre Drücke mit beträchtlichen Werten. Im ersten Fall wird Öl oder Quecksilber aus einem Vorratsgefäß angesaugt und durch ein Ventil in den Kompressionsraum gedrückt (*Cailletet-Pumpe*, Abb. 7.1.2–1), wodurch Drücke von 200 bis 500 MPa erzeugt werden. Bei der Druckschraube wird dagegen der sorgfältig eingeschliffene Kolben durch eine Schraubenspindel gegen die Kompressionsflüssigkeit gedrückt. Bei den hydraulischen Pressen wird ein axialer Doppelkolben mit zwei verschiedenen Querschnitten benutzt. Nach Erzeugen eines Vordruckes auf der Hochdruckseite mit einer Cailletet-Pumpe wird diese auf die Niederdruckseite des Doppelkolbens geschaltet, der als Druckübersetzer wirkt. Der gewünschte Druck ergibt sich durch Multiplikation des dort erzeugten Druckes mit dem Verhältnis der Kolbenflächen. Gasblasen in der Flüssigkeit würden wegen ihrer hohen Kompressibilität die Erreichung großer Drücke verhindern. Hydraulische Pressen bewähren sich von 300 bis 1 500 MPa. Bei höheren Drücken tritt ein Fließen auch der besten Stähle auf, weswegen man dann Sinterhartmetalle verwendet.

Sehr hohe Drücke zwischen ≈ 1 500 und 10 000 MPa lassen sich unter Verwendung spezieller, von BRIDGMAN angegebener Konstruktionen erreichen (Abb. 7.1.2–2). Hierbei wird auch die Kolbeninnenseite einem hohen Druck ausgesetzt und der Druckzylinder erfährt von außen durch eine konische Gestaltung einen zunehmenden Druck. Durch Sinterwerkstoffe und ausgewählte Stahlsorten sucht man das durch plastische Deformation auftretende große Kolbenspiel zu unterdrücken. Sehr hohe Drücke können auch dynamisch, d. h. für kurze Zeiten, mit Hilfe von Stoßwellen durch Explosivladungen oder ähnlich erzeugt werden.

Besondere Probleme bei hohen Drücken bereiten die verschiedenen erforderlichen Zugänge in das eigentliche Druckvolumen, die zur Durchführung von Experimenten nötig sind, vor allem isolierte elektrische Leitungen und optische Fenster. Geeignete Fenstermaterialien sind je nach dem erforderlichen Wellenlängenbereich Glas, Quarz, NaCl, CaF_2, Saphir und Silicium. Für kleine Volumina wurden Diamantzellen entwickelt.

7.1.3. Vakuumerzeugung

7.1.3.1. Laboranwendungen des Vakuums. Üblicherweise wird von einem Vakuum gesprochen, wenn in einem Behältnis oder Rezipienten der Druck der eingeschlossenen gas- oder dampfförmigen Stoffe durch „Evakuieren" mittels einer oder mehrerer „Vakuumpumpen" erheblich unter den Atmosphärendruck herabgesetzt worden ist. Die Anwendung solcher Vakua im Labor (und in der Technik) ist sehr vielfältig, wobei eine Druckskala von 15 Zehnerpotenzen überstrichen werden kann. Man unterteilt daher den Anwendungsbereich in Vorvakuum (Atmosphärendruck bis etwa 10^{-2} Pa, auch weiter unterschieden in Grob-, Zwischen- und Feinvakuum, in Hochvakuum (etwa 10^{-2} bis 10^{-5} Pa) und schließlich in Ultrahochvakuum (ab etwa 10^{-6} Pa). Der letzte Bereich kann noch in Höchst- und Extremvakuum (ab 10^{-8} Pa) untergliedert werden; oft wird im Hinblick auf die Erzeugung auch von einem treibmittelfreien Vakuum gesprochen.

Für Meßzwecke wird (Ultrahoch-)Vakuum besonders zur Untersuchung sauberer Oberflächen sowie von Prozessen des Adsorption, Oberflächendiffusion, katalytischen Wirksamkeit sowie der Elektronenemission eingesetzt. Verschiedene Verfahren der Elektronenspektroskopie werden verwendet zur Charakterisierung der Zusammensetzung (chemische Analyse durch Auger-Spektroskopie und Röntgen-Fotoeffekt), der elektronischen Struktur (UV-Fotoeffekt und Elektronenenergieverlustspektroskopie), der geometrischen Kristallstruktur (Beugung langsamer Elektronen) und der atomaren Schwingungen (hochauflösende Verlustspektroskopie) sowohl reiner als auch adsorbatbedeckter Oberflächen. In Anlagen zur Weltraumsimulation dient das Vakuum zur Materialprüfung unter den Bedingungen kleiner Drücke und extremer Temperaturen. Ähnliche Anwendungen liegen bei Tieftemperaturexperimenten unter Benutzung von Kryostaten vor. Viele Geräte und Anlagen, die mit elementaren Partikeln arbeiten, erfordern mindestens Hochvakuum, z. B. Elektronenmikroskope oder Ionenstrahlgeräte zur Materialanalyse und Ionenimplantation. Auch in den großen Beschleunigeranlagen der Kern- und Elementarteilchenphysik wird heute häufig Ultrahochvakuum gebraucht und erzeugt. Bei der Anwendung von Gasentladungsplasmen, die in Tokamakanlagen zur Entwicklung der Energiegewinnung durch Kernfusion gipfelt, wird zwar kein sehr niedriger Druck, dafür aber extreme Reinheit im Reaktionsraum verlangt, d. h. verschwindend kleine Partialdrücke störender Komponenten. Wichtig sind Vakuumverfahren im Labor auch bei der Herstellung „Dünner Schichten" verschiedenster Art sowie von Kontakten, die hauptsächlich durch Aufdampfen, aber auch durch Katodenzerstäubung aufgebracht werden. Zur Verbesserung bestimmter Materialeigenschaften werden Metalle im Vakuum durch Hochfrequenzglühen entgast oder extrem

dichte Verbindungen zwischen Metallen (und Keramik) durch Vakuumlöten und -schweißen erzeugt. Schließlich stellt auch die Reproduzierung der Druckskale für kleine Drücke nicht unerhebliche Anforderungen an die Laborvakuumtechnik.

7.1.3.2. Pumpen zur Vakuumerzeugung. Von den zahlreichen Pumpenkonstruktionen können hier nur die gebräuchlichsten erwähnt werden. Im Vorvakuumbereich werden häufig *Drehschieber-Vakuumpumpen* verwendet, die nach dem Verdrängerprinzip arbeiten (Abb. 7.1.3–1). Die angesaugten Gase und Dämpfe werden in einem (beweglichen) Raum mit sichelförmigem Querschnitt komprimiert und durch ein Ventil in die Atmosphäre befördert, was durch die exzentrische Lagerung des Rotors sowie die darin federnd gelagerten Schieber verwirklicht wird. Je nach ein- oder zweistufiger Ausführung werden für permanente Gase Enddrücke um 10^{-1} bzw. 10^{-2} Pa erreicht. Dämpfe können durch Anwendung des sogenannten Gasballastes abgepumpt werden, wodurch ihre Kondensation im Förderraum unterbunden wird. Für übliche Laboraufgaben werden i. allg. Pumpen bis ≈ 4 m³/h Saugvermögen eingesetzt.
Für Ultrahochvakuumapparaturen werden als Vorpumpen oft *Adsorptionspumpen* verwendet, die nach dem Speicherprinzip die abzupumpenden Gase an der Oberfläche bzw.

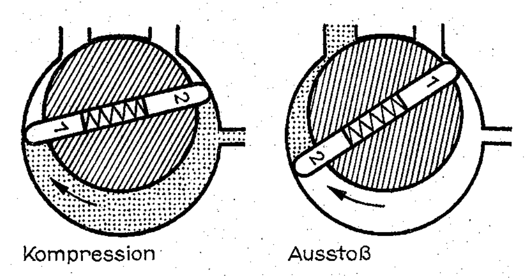


Abb. 7.1.3–1 Drehschieber-Vorvakuumpumpe. Der exzentrisch gelagerte Rotor wird mittels Elektromotor angetrieben, die Drehschieber sind federnd gelagert

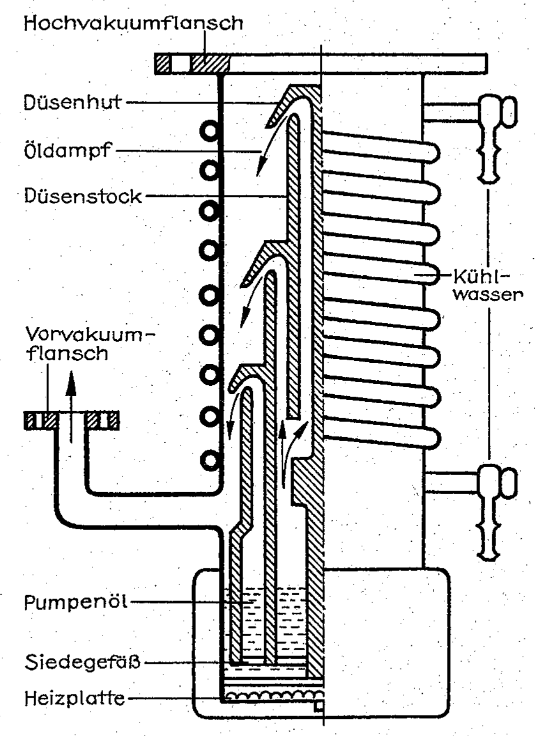


Abb. 7.1.3–2 Dreistufige Öldiffusionspumpe mit Wasserkühlung für Hochvakuumerzeugung

in den Poren von Molekularsieben binden (z. B. Zeolith-Molekularsieb 5A), solange sie sich auf tiefen Temperaturen befinden (Kühlung mit flüssigem Stickstoff). Da der Gleichgewichtsdruck über dem Adsorbens von der Beladung und der Gasart abhängt, ist das erzielbare Endvakuum etwas unterschiedlich, doch werden Drücke im 1-Pa-Bereich und eventuell auch darunter routinemäßig erreicht. Beim Erwärmen auf Zimmertemperatur entsteht ein Überdruck, der sich über ein Sicherheitsventil ausgleichen muß.
Im Hochvakuumbereich werden häufig *Diffusionspumpen* eingesetzt, bei denen der mittels Siedegefäß und Düsenstock erzeugte Strahl eines Treibmittels die abzupumpenden Gase mitreißt, die durch Diffusion in ihn eindringen (Abb. 7.1.3–2). Die Dampfströmung erlangt durch geeignete Düsenformen Überschallgeschwindigkeit, auf der Vorvakuumseite muß dann der Dampf an einer gekühlten Fläche wieder kondensieren und so ins Siedegefäß zurückkommen, weswegen Diffusionspumpen stets mit einer Wasserkühlung zu betreiben sind. Durch weiterentwickelte Pumpenöle wurde das giftige Quecksilber als Treibmittel weitgehend verdrängt und wird im wesentlichen nur noch in gläsernen Diffusionspumpen und Apparaturen benutzt, die allerdings in Verbindung mit stickstoffgekühlten Hg-Dampfsperren ein sehr „sauberes" Ultrahochvakuum erreichen. Auch mit Öldiffusionspumpen müssen durch Wasser, Peltier-Kühlelemente oder flüssigen Stickstoff gekühlte Dampfsperren (Baffle) bzw. Tiefkühlfallen eingesetzt werden, um eine Treibmittelrückströmung in den Rezipienten zu verhindern.
Letzteres entfällt bei den *Molekularluftpumpen*, die bereits

von GAEDE vorgeschlagen wurden und heute in Form der sogenannten *Turbomolekularpumpe* verwendet werden. Unter Ausnutzung des Effektes der äußeren und inneren Reibung in Gasen werden diese durch einen Satz von geeignet gefrästen Rotor- und Statorscheiben gefördert, wobei Umdrehungszahlen von ≈ 1 600 U/min benutzt werden. Neben horizontal gelagerten Achsen mit Kugellagern im Vorvakuumteil wurden später auch vertikale Anordnungen, z.T. mit Luftlagern entwickelt. Diese Pumpen können ebenfalls bis in den Ultrahochvakuumbereich eingesetzt werden, erfordern aber größere Sorgfalt und Wartung als Diffusionspumpen.

Zur Ultrahochvakuumerzeugung haben sich weiterhin *Ionengetterpumpen* sehr bewährt, die nach dem Speicherprinzip arbeiten und als Dioden- oder Triodenpumpen, letztere mit besserem Edelgassaugvermögen, ausgeführt werden. Die elektrische Ionsierung der abzupumpenden Gase in einer Penning-Entladung führt zum Steckenbleiben der erzeugten Ionen in der Titaniumkatode (Ionenpumpe, für Edelgase allein wirksam). Dabei wird jedoch auch Titanium zerstäubt, das auf den Wänden und den Anodenblechen eine langsam wachsende Schicht bildet, die durch Chemisorption ein Festhalten (Gettern) der aktiven Gase bewirkt. Die erforderlichen Permanentmagnete können i. allg. mit der Pumpe zusammen ausgeheizt werden.

Schließlich sind noch *Kryopumpen* zu nennen, die als Kondensations- oder Adsorptionspumpen wirken können. Im ersten Falle ergibt sich der bestenfalls erreichbare minimale Enddruck als Dampfdruck bei der Temperatur der kalten Fläche. Der sogenannte Cryotrapping-Effekt kann aber bewirken, daß auf dem Kondensat schwer kondensierbare Gase gebunden werden, die entsprechend der Temperatur eigentlich nicht kondensieren würden. Auch andere Adsorbentien, wie Molekularsiebe, werden zur Erhöhung des Saugvermögens eingesetzt. Neben Badkryopumpen können auch sogenannte Verdampferpumpen mit ständiger (aber geregelter) Kühlmittelzufuhr benutzt werden. Wie die Dampfdruckkurven zeigen, ist für Kryopumpen i. allg. nur ein Betrieb mit flüssigem Helium sinnvoll, was wegen der Kosten und nötigen Rückgewinnung verschiedene Probleme mit sich bringt.

7.1.3.3. Technik der Vakuumerzeugung. Die Erreichung eines Vor- oder Hochvakuums ist ohne besondere Probleme möglich. Alle im Hochvakuumbereich arbeitenden Pumpen brauchen Vorpumpen zur Erzeugung des erforderlichen Vorvakuums bzw. zum Starten (Ionengetterpumpen), wofür je nach Gasanfall und sonstigen Anforderungen Drehschieber- oder Adsorptionspumpen eingesetzt werden. Die Verbindung erfolgt durch Vakuumschlauch und genormte Flansche, die mit Gummiringen gedichtet werden (Abb. 7.1.3–3). Bei Glasapparaturen werden anstelle der Flansche mit Vakuumfett gedichtete Schliffverbindungen verwendet. Auf der Hochvakuumseite können mit Vorteil Vitondichtringe eingesetzt werden, die bis 100 °C ausheizbar sind und einen Enddruck von 10^{-6} Pa ermöglichen. Zur Verbesserung des Vakuums in größeren Rezipientenglokken können Meißner-Kühlfallen mit flüssigem Stickstoff und eine Vorentgasung durch Ausheizen angewendet werden.

Zur Erzeugung von Ultrahochvakuum sind besondere Vorkehrungen zu treffen. Alle Dichtungen des Rezipienten bis hin zum Anschluß der Pumpe müssen entweder aus Viton oder (für Drücke unter 10^{-6} Pa) aus einem Metall bestehen, meist Kupfer, aber auch Gold oder Aluminium. Rezipienten, Rohrverbindungen usw. sowie die Flansche mit speziellen Schneidenformen werden aus Edelstahl gefertigt. Auch die Strom- und Kühlmittelzuführungen und die Ventile müssen ultrahochvakuumtüchtig sein. Für Ventile und die Übertragung von Bewegungen ins Vakuum werden Metallfaltenbälge eingeschweißt. Metallapparaturen müssen zur Entfernung von Wasserhäuten u. a. Wandverunreinigungen stundenlang bis 250 °C ausgeheizt werden, Glasapparaturen sogar bis 400 °C. Günstig ist es, auch die Pumpen auszuheizen, was bei Ionengetter- und Getterpumpen ohne weiteres möglich ist. Besonderes Augenmerk ist auf die Dichtigkeit der Bauteile und Flanschverbindungen zu legen. Die Lecksuche spielt daher eine große Rolle. Nach Beseitigung der (unter Umständen erst beim Ausheizen entstandenen) Lecks ist die Erzeugung des Ultrahochvakuums weniger ein Problem der Gasförderung, sondern der Beherrschung der Gasabgabe von den Wänden, Elektroden, Glühkatoden usw. Deswegen sind neben den Ausheizprogrammen auch oft noch spezielle Ausglühzyklen nötig, auch ist eine sorgfältige Entgasung der Meßröhren erforderlich, besonders wenn heiße Katoden benutzt werden. Schwierigkeiten treten auf, wenn Dämpfe oder Gase eingelassen werden müssen oder ständig zuströmen, z. B. beim Betrieb fensterloser Gasentladungslampen zur Erzeugung von Licht im Vakuum-UV-Bereich zur Fotoelektronenspektroskopie.

Zur Lecksuche werden sowohl eigens dafür entwickelte Geräte als auch die Total- und Partialdruckmeßgeräte der vorliegenden Apparaturen selbst verwendet. Zweckmäßigerweise wird zunächst nach Abschalten der Pumpen durch

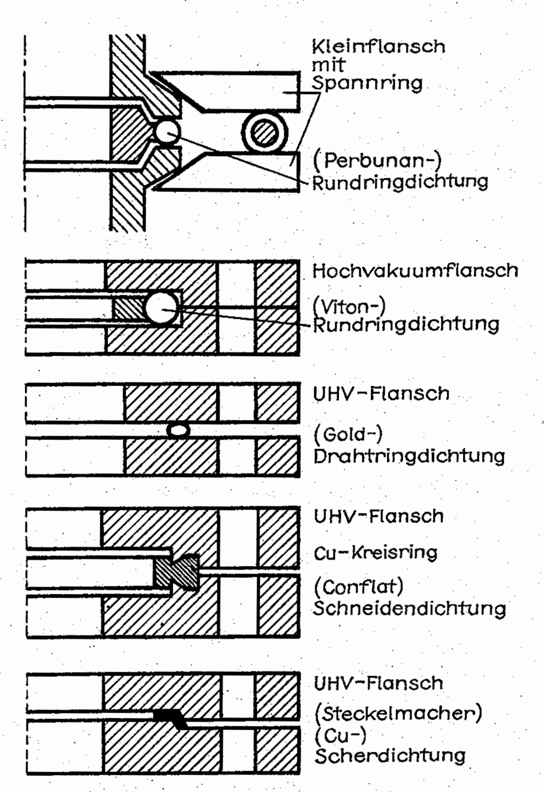


Abb. 7.1.3–3 Flanschverbindungen der Vakuumtechnik für Vorvakuum-, Hochvakuum- und Ultrahochvakuumzwecke

eine Druckanstiegsmessung die Größenordnung der Leckrate bestimmt. Durch Besprühen oder Abpinseln mit einem geeigneten Stoff sucht man die Stellen der im Betrieb befindlichen Apparatur, an denen durch Lecks das Testgas o. ä. eindringt und die Partialdruckzusammensetzung im Vakuum ändert. Steht kein Massenspektrometer bzw. Lecksuchgerät zur Verfügung, kann die unterschiedliche Empfindlichkeit von Vakuummetern für verschiedene Gase oder Dämpfe ausgenutzt werden, wie auch der Effekt der Verstopfung eines Lecks mit organischen Lösungsmitteln.
Die relativ einfachen Halogenlecksuchgeräte arbeiten mit gasförmigen Halogenkohlenwasserstoffen, die an einer Platinglühanode mit Erdalkalizusatz eine Änderung des Ionenstromes hervorrufen. Am günstigsten und empfindlichsten sind Heliumlecksuchgeräte, die aus einem kleinen Massenspektrometer bestehen, das i. allg. fest auf Helium eingestellt ist. Dieses ist als Testgas nicht nur ungefährlich, sondern eignet sich gut wegen der Kleinheit und Reaktionsträgheit der Atome, wobei noch Lecks bis herab zu 10^{-11} Pa l s^{-1} entdeckt werden können. Auch sind Eichlecks erhältlich, mit denen die Empfindlichkeit eingestellt oder überprüft werden kann. Bei ungünstiger Materialauswahl oder Konstruktion können übrigens sogenannte virtuelle Lecks auftreten, die ein reelles Leck vortäuschen, indem sie z.B. über Poren oder Haarrisse für lange Zeiten Gase abgeben. Das ist besonders für Anlagen von Bedeutung, wo Leckraten von $< 10^{-7}$ Pa l s^{-1} gefordert werden, während für Hochvakuumanlagen die restlichen Lecks $< 10^{-3}$ Pa l s^{-1} sein sollten.

7.1.4. Erzeugung hoher elektrischer und magnetischer Felder

7.1.4.1. Laboranwendungen hoher Felder. Elektrische und magnetische Feldstärken werden normalerweise nicht nach allgemein akzeptierten Bereichen eingeteilt, doch wird man elektrische Felder ab 10^7 Vm^{-1} und magnetische Felder ab ≈ 2 T als groß bezeichnen können. In Elektromagneten erzeugte Felder werden üblicherweise in Tesla, also der Einheit der magnetischen Induktion (Vs/m^2), gemessen (früher im cgs-System in Gauß entsprechend 10^{-4} T), die elektrische Feldstärke wird häufig auch in Volt je Zentimeter angegeben. Die Anwendung mittlerer und hoher elektrischer und magnetischer Felder im Labor ist sehr umfangreich, so daß nur einige Beispiele herausgegriffen werden können. Elektrische Felder werden besonders zur Beschleunigung geladener Teilchen eingesetzt, z. B. von Elektronen, Protonen oder Ionen (auf die verschiedenen Beschleuniger wird an anderer Stelle eingegangen). Sie werden weiterhin zur Anregung und Ionisierung von Molekülen und Atomen verwendet, wobei im ersten Fall der Stark-Effekt zur Aufspaltung und Verbreiterung von Spektrallinien führt. Die Feldionisation wird zur Abbildung der atomaren Struktur von Festkörperoberflächen und einzelner Atome im Feldionenmikroskop nach E.W. Müller sowie zur Massenspektroskopie angewendet.
Elektrische und magnetische Felder werden in mannigfachen Kombinationen zur Ablenkung und Bündelung bzw. Fokussierung von Teilchenstrahlen in Beschleunigern und Speicherringen wie auch zum Zusammenhalt von Gasentladungsplasmen, z.B. in Tokamak-Fusionsexperimenten, eingesetzt. Weit verbreitet ist auch ihre Ausnutzung zu Abbildungszwecken in konventionellen Durchstrahlungs- bis hin zu Höchstspannungselektronenmikroskopen sowie in Rastermikroskopen mit ihren vielfältigen Nachweis- und Abbildungsmoden. Der Einsatz von Materialien für solche Anordnungen wie auch in der (Hochspannungs-)Technik erfordert die Bestimmung der Feldstärkeabhängigkeit verschiedener Festkörpereigenschaften.
Hohe Magnetfelder dienen besonders zur Untersuchung der magnetischen und magnetooptischen Eigenschaften der Materie sowie auch zur Erzeugung sehr tiefer Temperaturen. Ein breites Anwendungsgebiet haben sie auch für analytische und physikalische Zwecke in Meßverfahren gefunden, die auf Resonanzeffekten beruhen, wie Elektronen- und Kernspinresonanz, Zyklotronresonanz, De-Haas-van-Alphen-Effekt sowie zur Massenspektrometrie und -separation.

7.1.4.2. Erzeugung hoher elektrischer Felder. Da sich die Feldstärke in einer ebenen Anordnung von Elektroden, z.B. im Plattenkondensator ohne Dielektrikum, aus dem Verhältnis von Spannung zu Abstand ergibt, entstehen hohe Felder stets beim Anlegen großer Spannungen an kondensatorähnliche Elektrodenanordnungen. Dabei kann es sich um Gleich- oder Wechselspannungen handeln, wobei letztere im Netzfrequenzbereich recht einfach durch Hochtransformation erzeugt werden, eventuell mit Öl- oder Epoxidharz(verguß)isolation der Transformatoren. Mit höherer Frequenz muß ein größerer Aufwand getrieben und schließlich auf Magnetron- oder Klystronsenderöhren zurückgegriffen werden. Für die Gleichspannungs- und Gleichfelderzeugung eignen sich für spezielle Zwecke Van-de-Graff-Generatoren, die eine mechanische Ladungstrennung durch ein bewegliches Band bewirken und im Prinzip mit den Reibungselektrisiermaschinen verwandt sind. Hierbei erhält man Spannungen bis zu mehreren Megavolt, wobei für die höheren Spannungen Drucktankgeneratoren mit einem Füllgas (Stickstoff) eingesetzt werden. Im Kilovoltbereich werden hohe Spannungen bequemer mit Netzgleichrichtern erzeugt. Neben eventuellen Isolationsproblemen bei der Hochtransformation der Wechselspannung ist zu beachten, daß beim Einsatz von Gleichrichterröhren eine auf Hochspannungspotential liegende Heizspannung erforderlich ist, vor allem bei Spannungsvervielfacherschaltungen (Greinacher-Schaltung). Werden hohe elektrische Felder zur Teilchenbeschleunigung benutzt, z.B. für Ionen, müssen oft erhebliche Teile der Apparatur auf einem hohen elektrischen Betriebspotential gegen Erde liegen, nämlich die Ionenquelle mit ihrer Glühkatode, das Gaseinlaßsystem, die Kühlwasserzuführung sowie Strom- und Spannungsregelvorrichtungen und die zugehörigen Meßgeräte.
Die Realisierung hoher Felder erfordert besondere Vorkehrungen hinsichtlich der Elektrodengestaltung, Oberflächenbeschaffenheit und Sauberkeit zur Vermeidung von Überschlägen (Felddurchschlag) und bezüglich der eingesetzten Isolatoren. Für Laborzwecke eignen sich neben reiner Al_2O_3-Keramik, Glas, Porzellan und Quarz auch eine bearbeitbare Keramik, bei Glas stört allerdings oft eine leitende Oberflächenschicht sowie eine Wasserhaut. Außer dem spezifischen Widerstand (Quarz 10^{17} bis 10^{18} Ωcm) sind die Durchschlagsfestigkeit (Quarz 100 bis 150 kV/cm, Porzellan um 350 kV/cm) und für Hochfrequenzspannungen auch die elektrischen Verluste von Bedeutung. Für Hochspannungszuführungen ins Vakuum wurden besondere Technologien der Metall-Keramik-Verbindung entwickelt. Hohe Feldstärken werden auch erreicht, wenn der Abstand der Elektroden sehr verringert wird. Das ist meist nur mit Hilfe einer geeigneten dünnen Isolatorschicht möglich, z. B. für die Untersuchung des Feldeffektes bei Halbleitern bzw. sei-

ner Anwendung in Bauelementen der Mikroelektronik (MOSFET). Im ersten Falle lassen sich dünne Kunststofffolien oder Glimmerplättchen verwenden. Dünne Isolatorschichten lassen sich durch Aufdampfen, Ionenzerstäubung oder Oxidation herstellen. Die erreichbaren Feldstärken liegen bei einigen 10^7 V/m. Schließlich werden sehr hohe Feldstärken in Spezialfällen durch Geometrieeffekte erzielt: Durch Anätzen und thermisches Verrunden kann man die Enden von Drähten (teilweise auch von Halbleiterkristallen) mit einem schlanken Konus und einer abschließenden Spitze versehen, so daß dort Krümmungsradien ab einigen 10 nm entstehen. Bei der Anwendung und Beobachtung solcher Emitter im Feldelektronen- oder Feldionenmikroskop werden beim Anlegen von Spannungen im kV- oder 10-kV-Bereich Feldstärken über 10^9 V/m bzw. 10^{10} V/m erzeugt.

7.1.4.3. Erzeugung hoher Magnetfelder. Magnetfelder bis zu mittleren Größen werden in konventionellen Elektromagneten erzeugt, bei denen jedoch die Sättigungsmagnetisierung des Polschuhmateriales eine Grenze setzt. Je nach dem erforderlichen Feldvolumen, den Homogenitätsforderungen und dem verwendeten Material können durch geeignete Polschuhformen Felder bis etwa 5 T erreicht werden. Dabei ist die elektrische Verlustleistung durch Kühlwasser abzuführen, am günstigsten in Kupferrohren, die zugleich die Wicklung bilden. Größere Felder lassen sich in sogenannten Bittermagneten herstellen, die aus zentrisch durchbohrten Kupferscheiben bestehen und mit einer Zwischenisolation zusammengepreßt werden, aber nach Art einer gestauchten Wendeltreppe miteinander verbunden sind. Neben dem stationären Betrieb mit großen Strömen werden solche Magnete besonders zur Erzeugung von sehr hohen Impulsmagnetfeldern benutzt, indem ein Gleichstromgenerator (Dynamomaschine) oder ein Wechselstromgenerator in einer Halbperiode bzw. eine aufgeladene Kondensatorbatterie über den Magneten kurzgeschlossen wird (KAPITZA). Kurzzeitig sind Felder von mehr als 1 kT erreichbar. In jüngerer Zeit werden supraleitende Magnete eingesetzt, die sich in einem Kryostaten mit flüssigem Helium befinden, doch besteht eine prinzipielle Grenze durch das Verschwinden der Supraleitung bei zu hohen Feldern. Zur Stromleitung werden mit Kupfer oder Aluminium überzogene supraleitende Drähte u. a. aus Niobium-Titanium-Legierungen verwendet, so daß der Normalleiter den Strom bei höheren Temperaturen übernimmt. Die gespeicherte magnetische Energie erreicht beträchtliche Werte und muß beim Abschalten in geeigneter Weise abgeführt werden, damit keine Zerstörungen auftreten. In käuflichen Magneten werden 8 T ohne Schwierigkeiten erzeugt, Laborausführungen erreichen auch wesentlich höhere Werte, z.B. 14 T bzw. 11,4 mA/m in einer Spule von 600 kg, oder werden für besonders große Abmessungen oder spezielle Magnetfeldformen konstruiert (magnetische Flasche für Plasmaexperimente).

7.1.5. Erzeugung und Messung von Teilchenstrahlen

7.1.5.1. Laboranwendungen von Teilchenstrahlen. Unter diesem Begriff sollen gebündelte Teilchenströme von Elektronen, Ionen, Atomen oder Molekülen verstanden werden, die sich i. allg. im Vakuum befinden und für verschiedene Zwecke im Labor genutzt werden. Neben der bekanntesten Anwendung im Elektronenmikroskop zu Abbildungszwekken werden Elektronenstrahlen auch zur Elektronenbeugung herangezogen, meist als wählbare zusätzliche Betriebsart in Durchstrahlung oder bei streifendem Einfall. In Rasterelektronenmikroskopen können neben der eigentlichen Abbildung auch verschiedene Analysenbetriebsarten angewendet werden. In Oberflächenanalysengeräten werden Elektronen zur Bestimmung von Oberflächen- und Adsorbatstrukturen unter Ultrahochvakuum-Bedingungen genutzt, und zwar langsame Elektronen (50 bis 500 eV) meist bei senkrechtem Einfall (LEED) und mittelschnelle Elektronen (25 bis 50 keV) bei streifender Inzidenz (RHEED). Die Auger-Elektronenspektroskopie (AES) wird als oberflächenempfindliche Analysenmethode eingesetzt und die Elektronenverlustspektroskopie (ELS) in Transmissions- und Reflexionsmode zur Aufklärung elektronischer Volumen- und Oberflächeneigenschaften. Schließlich werden Elektronenstrahlen zur Oberflächenreinigung herangezogen und künftig bei der Elektronenstrahllithographie eine zunehmende Bedeutung für die Herstellung kleinster Strukturen in der Mikroelektronik gewinnen.

Ionenstrahlen werden ebenfalls für die chemische Analyse eingesetzt (bei Energien von 1 MeV z.B. zur Rutherford-Streuung oder bei kleineren Energien in der Sekundärionenmassenspektroskopie, kurz SIMS) wie auch für die Aufklärung von Oberflächen- und Adsorptionsstrukturen (ISS). Eine wichtige Rolle spielen Ionenstrahlen in der Halbleitertechnologie für die Dotierung mittels Ionenimplantation, wodurch sich scharfe Dotierungsprofile erreichen lassen.

Atom- und Molekularstrahlen werden z. B. für die Untersuchung der optischen Anregung oder Fotoionisation in Gasen oder der Adsorptionskinetik an Oberflächen eingesetzt. In jüngerer Zeit ist mit der sogenannten Molekularstrahlepitaxie eine Technik entwickelt worden, die unter Ultrahochvakuumbedingungen Halbleiter mit vorgegebenen Schichtstrukturen durch Aufdampfen aus gesteuerten Knudsen-Zellen herzustellen gestattet.

7.1.5.2. Erzeugung von Teilchenstrahlen. Elektronenstrahlen werden in sogenannten Elektronenkanonen erzeugt, wie sie auch in Oszillografen- und Fernsehbildröhren verwendet werden. Sie bestehen aus der elektronenemittierenden Katode, dem Wehnelt-Zylinder mit einem gegenüber dem Emitter negativen Potential und (mindestens) einer Anode mit einem positiveren Beschleunigungspotential. Während für kommerzielle Zwecke meist (Erdalkali-)Oxidkatoden eingesetzt werden, bewähren sich im Labor die Wolframdrahtkatoden wegen ihrer Unempfindlichkeit gegenüber wiederholten Belüftungen. Die erforderliche Temperatur für eine genügend starke Elektronenemission ist allerdings recht hoch ($\approx$ 2 200 °C für 100 mA/cm²), weswegen teilweise andere Katoden für niedrige Temperaturen bevorzugt werden, z. B. thoriertes oder karburiertes Wolfram, Iridium oder Rhenium mit ThO_2- oder Y_2O_3-Überzug oder Lanthanhexaborid (LaB_6). In speziellen Fällen werden für größte elektronenoptische Auflösung auch Feldemissions(kalt)katoden benutzt, die nahezu punktförmige Emission bei sehr hohen Stromdichten ergeben, aber Ultrahochvakuum-Betriebsbedingungen erfordern.

Ionen werden in vielen Fällen durch Stoßprozesse von Elektronen erzeugt, die durch eine Glühkatode emittiert werden, wobei allerdings die erreichbaren Stromstärken der mit einer Ionenoptik formierten Ionenstrahlen klein sind (μA-Bereich). Intensivere Strahlen erhält man durch Bogenentladungen in Gasen mit vermindertem Druck. Plasmatrons und Duoplasmatrons liefern hohe Ionenströme im mA-Bereich, doch muß der relativ hohe Gasbetriebsdruck in mehreren Druckstufen durch sogenanntes differentielles

Pumpen herabgesetzt werden, um den durch die Kapillaren oder Blenden gefädelten Ionenstrahl im Hochvakuum- oder Ultrahochvakuumbereich zu nutzen.
Die Strahlen geladener Teilchen können durch elektrische Felder bequem nachbeschleunigt werden, eventuell in mehreren Stufen oder Beschleunigungsstrecken. Außerdem lassen sich Monochromatoren einsetzen, die ihre Geschwindigkeits- oder Energieverteilungsbreite schmaler machen, z. B. bei Elektronen für hochauflösende Analyse. Bei Ionen können durch magnetische Sektorfelder störende Fremdionen aus dem Strahl ausgelenkt und damit entfernt werden.
Durch Verdampfung erzeugte Teilchen„strahlen" haben ohne besondere Vorkehrungen i. allg. eine kosinusförmige Flugrichtungsverteilung, wie sie zur Schichtherstellung oder Beschichtung größerer Objekte verwendbar ist. Die Verdampfungsquellen haben mannigfache Formen und können durch Stromdurchgang widerstandsbeheizt, durch eine Glühwendel indirekt erhitzt, durch Elektronenstoß oder Hochfrequenz oder sogar durch einen Laserstrahl aufgeheizt werden. Auch verschiedene Formen der Katodenzerstäubung können zur Schichtherstellung benutzt werden, wobei die Katode aus dem aufzubringenden Material besteht. Dabei werden Ionenstrahl- und Plasmazerstäuben in verschiedenen Abwandlungen und Kombinationen, wie reaktives oder ionenstrahlgestütztes Plasmazerstäuben, eingesetzt, womit sich auch kaum verdampfbare Verbindungen, z. B. für Isolationsschichten, handhaben lassen. Für eine besonders definierte und steuerbare Schichtherstellung, wie sie in der Molekularstrahlepitaxie (MBE) erforderlich ist, eignen sich Knudsen-Zellen, d. h. Atomstrahlöfen mit mehreren Blenden. Um den Druckanstieg im Rezipienten während der Bedampfung weitgehend zu unterbinden, werden mit flüssigem Stickstoff gekühlte Strahlungsschirme eingesetzt. In Ultrahochvakuum-MBE-Anlagen zur Herstellung von Halbleitern mit Überstrukturen verschiedener, oft periodischer Materialzusammensetzung in atomaren Dimensionen werden die Knudsen-Zellen computergesteuert.

7.1.5.3. Messung von Teilchenstrahlen. Die Bestimmung der Intensität von Teilchenstrahlen wird meist auf eine Strommessung oder eine Einzelteilchenzählung zurückgeführt. Geladene Teilchen, wie Elektronen und Ionen, werden in einem Faraday-Becher aufgefangen, einem zylindrischen Gefäß mit kleiner axialer Öffnung, vor der sich noch eine Ringblende befindet mit einem Potential, welches das Entweichen der entstehenden Sekundärelektronen verhindert. Neutrale Atome oder Moleküle können in einer vorgeschalteten Ionisierungsstrecke durch Elektronenstoß ebenfalls geladen werden, wonach sich in allen Fällen der Teilchenstrom als elektrischer Strom mit Elektrometern o. a. Strommessern bestimmen läßt. Bei besonders schwachen Strömen sowie zur Einzelteilchenzählung werden Sekundärelektronenvervielfacher (SEV) dem eigentlichen Verstärker vorgeschaltet, um das Signal-Rauschverhältnis zu erhöhen. SEV bestehen oft aus CuBe-Blechen in Form offener Kästchen (Dynoden), die unter Ausnutzung der Sekundärelektronenemission eine 10^5– bis 10^7fache Verstärkung der auftreffenden Teilchenströme bewirken. Da die Stufenspannungen 300 bis 400 V betragen und meist 12 bis 14 Stufen verwendet werden, liegt die Ausgangselektrode (Anode) auf einem Potential von 3 bis 4 kV gegenüber der ersten Dynode, woraus Probleme beim Übergang zum Vorverstärker resultieren, dessen Ausgang ungefähr auf Erdpotential liegen möchte. Ionen werden zweckmäßigerweise auf eine Prallplatte beschleunigt, wo Elektronen ausgelöst und mit einem SEV verstärkt werden. Anstelle dieser offenen SEV werden auch Kanalvervielfacher verschiedener Konstruktion eingesetzt, die z. B. als Channeltrons wesentlich kleiner sind, aber ebenfalls eine Spannung von einigen Kilovolt brauchen. Letztere bestehen aus einem spiralförmigen Glasröhrchen, das innen einen halbleitenden Belag mit großem Sekundäremissionskoeffizienten enthält, so daß durch die anliegende Hochspannung eine lawinenartige Verstärkung des Elektronenstromes erfolgt (Abb. 7.1.5–1). Denkt man sich diese Röhrchen weiter verkleinert und etwa bienenwabenartig in großer Anzahl nebeneinander verschmolzen, entstehen sogenannte Channelplates oder Kanalplattenverstärker, mit denen z. B. im Feldionenmikroskop direkt ein zweidimensionales Ionenbild erzeugt wird. Verschiedene Schwierigkeiten mit dem innerhalb des Rezipienten angeordneten (offenen) SEV umgeht man, wenn dort lediglich ein Leuchtschirm oder Szintillatorkristall angebracht wird, die beim Auftreffen der Teilchen entstehenden Lichtblitze aber durch ein Fenster mit einem abgeschmolzenen Fotovervielfacher (SEV mit Fotokatode) registriert werden. Stark ionisierende Teilchen lassen sich auch mit Zählrohren nachweisen, die als „Geigerzähler" funktionieren, indem in einem gasgefüllten Rohr durch ein starkes elektrisches Feld sekundär erzeugte Ionen beschleunigt werden, wodurch eine als Strom nachweisbare Ionenlawine ausgelöst wird.
Neutrale Moleküle können wie erwähnt durch Stoß ionisiert und dann nachgewiesen werden. Interessiert die Teilchenmasse, werden die betreffenden Molekularstrahlen in die Ionenquelle eines Massenspektrometers geleitet, also z. B. in eine Geradsichtvariante eines Quadrupolmassenspektrometers, womit das Verhältnis von Masse und Ladung bestimmt werden kann.
Oft wird auch die erzeugte Schichtdicke auf einem bedampften Substrat gebraucht. In Schichtdickenmeßgeräten wird dazu die Frequenzänderung eines gleichzeitig mit dem Substrat bedampften Schwingquarzes bestimmt, wodurch eine kontinuierliche Messung und Regelung ermöglicht wird. Auch die Änderung der optischen Absorption während der Bedampfung oder nachträglich durchgeführte Interferenzmessungen oder ellipsometrische Bestimmungen können ausgenutzt werden.

7.1.6. Erzeugung und Anwendung elektromagnetischer Strahlung

7.1.6.1. Anwendung elektromagnetischer Strahlung im Labor. Aus dem breiten elektromagnetischen Spektrum wird besonders die Licht- und Röntgen-Strahlung im Labor verwendet. Neben den „klassischen" Lichtquellen werden in neuerer Zeit auch Laser eingesetzt, die an anderer Stelle besprochen werden (vgl. 4.11.1.). Die Lichtstrahlung dient in einem großen Wellenlängenbereich vom Infrarot über das sichtbare Licht bis zum Ultraviolett und dem anschließenden Vakuumultraviolettbereich zur Bestimmung optischer Eigenschaften der Materie. Dazu gehören vor allem Reflexions- und Absorptionsvermögen sowie die damit im Zusammenhang stehenden (atomaren und eventuell auch kollektiven) Anregungen von Atomen, Molekülen und festen Körpern. Zur Auslösung von Fotoelektronen (UPS) wird besonders das Vakuum-UV-Licht ($\lambda < 200$ nm) verwendet, wobei aus deren Energieanalyse wichtige Aussagen zur Bandstruktur, chemischen Zusammensetzung und Bindung sowie zu elektronischen Oberflächeneigenschaften ge-

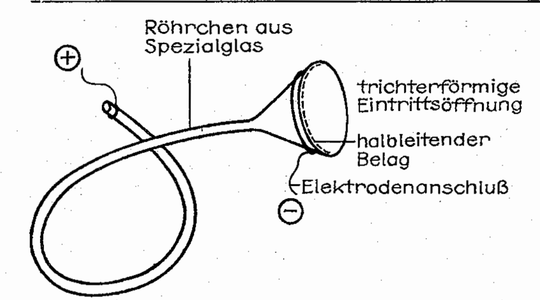


Abb. 7.1.5–1 Channeltron (Verstärkeranordnung im Vakuum für Elektronen, Spannung 3 bis 4 kV), betreibbar in Strommessungs- oder Impulszählschaltung

wonnen werden. Röntgen-Strahlen werden in großem Umfang zur Bestimmung von Kristallstrukturen und Strukturdefekten eingesetzt, aber auch zur Untersuchung der Absorption und Röntgen-Strahlenemission (Röntgen-Fluoreszenzanalyse). Da Röntgen-Strahlen auch Fotoelektronen auslösen, werden sie in sogenannten ESCA-Spektrometern zur Elektronenspektroskopie für die chemische Analyse benutzt. Schließlich ist die γ-Spektroskopie als ein wichtiges Analysenverfahren zu erwähnen.
Auch die elektromagnetische Strahlung der längerwelligen Seite des Spektrums wird für hochempfindliche und aussagekräftige Analysenverfahren genutzt. Elektromagnetische Strahlung im Radiofrequenzbereich wird in der Kernresonanzspektroskopie (nmr) in verschiedenen Varianten in Kombination mit (mittelstarken) Magnetfeldern eingesetzt. Während hier der Nachweis über die Kernspins erfolgt, die bei geeigneten Frequenzen und Magnetfeldern in Resonanz geraten, wird elektromagnetische Strahlung im cm-Wellenbereich für die Anregung von Elektronenspins in der Elektronenspinresonanzspektroskopie (epr), ebenfalls in Kombination mit einem Magnetfeld, verwendet.

7.1.6.2. Erzeugung von elektromagnetischer Strahlung. Obgleich sich für die Erzeugung von Licht im sichtbaren Bereich im Prinzip die üblichen Lichtquellen eignen, kann für Meßzwecke, z. B. bei Verwendung von Monochromatoren, die Wendelung des Leuchtdrahtes stören oder die Leuchtdichte zu gering sein. Im ersten Fall lassen sich Wolframbandlampen einsetzen, deren flächenhaft leuchtendes Wolframblech günstig auf den Eintrittsspalt abgebildet werden kann, auch stehen meist Eichkurven zur Verfügung. Die Helligkeit ist allerdings wesentlich geringer als die von Punktlicht- oder Halogenlampen. Vergleichweise große Lichtleistungen bei hoher Leuchtdichte haben die sogenannten Höchstdrucklampen mit Quecksilber- oder Xenonfüllung in einem hochbelastbaren Quarzkolben, bei denen Linienspektren dieser Elemente zusammen mit einem Kontinuum in einer Gasentladung angeregt werden. Der hohe Betriebsdruck von ≈ 6 MPa (60 atm) erfordert besondere Sorgfalt im Umgang sowie ein stabiles Gehäuse. Wegen der „fallenden" Kennlinie einer Gasentladung ist ein für einige Ampere dimensionierter Vorwiderstand erforderlich. Verwandt sind die Spektrallampen, die Linienspektren, wenngleich mit geringer Intensität ergeben, die zu Eichzwecken dienen. In den sogenannten Hohlkatodenlampen werden charakteristische Linien von Metalldämpfen erzeugt und für spektroskopische Zwecke (Atomspektroskopie) eingesetzt.
Laserlicht hat den Vorteil eines fast parallelen und intensiven Bündels und ist monochromatisch, abgesehen bei „Multimodenbetrieb" beispielsweise eines Argonlasers. Die Intensität ist oft so hoch, daß Schutzmaßnahmen erforderlich sind, da selbst an Teilen der Apparatur ungewollt reflektiertes Licht noch zu Augenschädigungen führen kann. Für spektroskopische Zwecke ist die Beschränkung auf wenige Wellenlängen nachteilig. Hier wurde mit dem in gewissen Bereichen „durchstimmbaren" Farbstofflaser eine Abhilfe geschaffen. Verschiedene Laser, die nur im Impulsbetrieb arbeiten wie der Stickstoff-Laser, erfordern bei ihrer Anwendung eine entsprechende und aufwendigere Impulsmeßtechnik.
Für das „nahe" Ultraviolett (bis ≈ 200 nm Wellenlänge) werden häufig die genannten Höchstdrucklampen eingesetzt, teilweise auch Deuteriumlampen. Im Infrarotbereich verwendet man außer den Höchstdrucklampen sogenannte Globare, die in ähnlicher Form als Silitstäbe zur Heizung von Laboröfen benutzt und durch direkten Stromdurchgang zum Glühen gebracht werden. Zur Monochromatisierung des Lichtes in den genannten drei Spektralbereichen dienen Prismen- oder Gittermonochromatoren. Im ersten Falle sind je nach Wellenlängenbereich und erforderlicher Auflösung Flintglas-, Quarz-, NaCl- o. a. Prismen auszuwählen. Gittermonochromatoren sind etwas lichtstärker, doch können sich in manchen Bereichen mehrere Ordnungen des

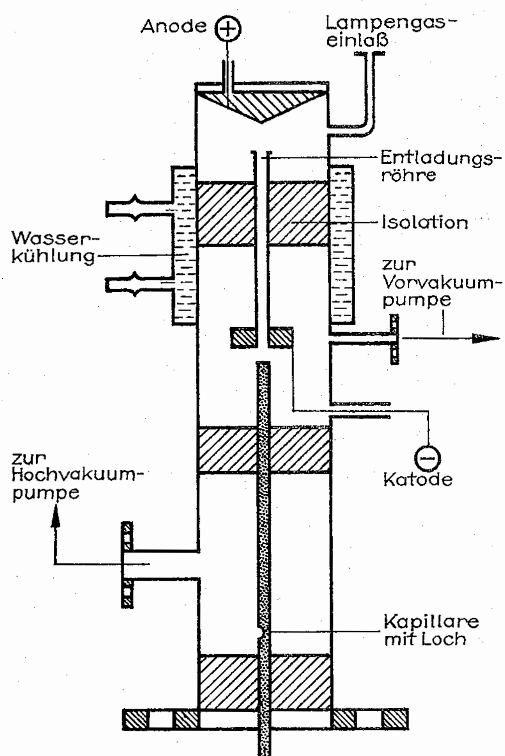


Abb. 7.1.6–1 Differentiell gepumpte Kaltkatoden-Gasentladungslampe für Vakuum-UV-Licht. Der nötige Brenndruck im Vorvakuumbereich und das Ultrahochvakuum werden mittels der im Hochvakuumbereich abgepumpten Kapillare realisiert

Spektrums überlagern, so daß unter Umständen eine Vorzerlegung erforderlich ist. In beiden Monochromatorarten werden Spiegeloptiken verwendet, wie auch die Gitter in Reflexionen eingesetzt werden.
Von ≈ 200 nm nach kürzeren Wellenlängen hin wird UV-Licht in Luft absorbiert, weswegen dieser Bereich als Vakuum-UV gekennzeichnet wird und die (speziell konstruierten) Monochromatoren evakuiert werden müssen. Lichtquellen sind wieder Gasentladungslampen, bei denen die Entladung in einem dünnen Quarzrohr brennt. Jenseits einer Lichtquantenenergie von ≈ 10 eV lassen sich keine Fenster mehr verwenden. Dann muß mit Hilfspumpen über Kapillaren ein Druckgefälle zum Monochromator (und eventuell zum Fotoelektronenspektrometer) hergestellt werden, wonach sich z. B. (He-)Linien von ≈ 20 und 40 eV für Fotoemissionsuntersuchungen im Ultrahochvakuum benutzen lassen (Abb. 7.1.6–1). Noch höhere Lichtenergien liefert die Synchrotronstrahlung, die bei der (Zentripetal-)Beschleunigung von Elektronen entsteht, die im Synchrotron auf einer Kreisbahn nahezu mit Lichtgeschwindigkeit umlaufen. Diese Strahlung erstreckt sich mit einem kontinuierlichen Spektrum vom Röntgen-Strahlen- bis zum Mikrowellenbereich und ist außerordentlich intensiv, stark gebündelt und polarisiert, allerdings auch zeitlich moduliert (beispielsweise treten beim Synchrotron NINA im Betrieb mit 1-GeV-Elektronen 25 kW Strahlungsleistung auf, wobei das Spektrum dem eines Schwarzen Strahlers von 10^7 K ähnelt, aber 1 bis 2 Größenordnungen intensiver ist). Mit Vakuummonochromatoren kann der Vakuum-UV-Bereich des Lichtes kontinuierlich überstrichen werden. Noch

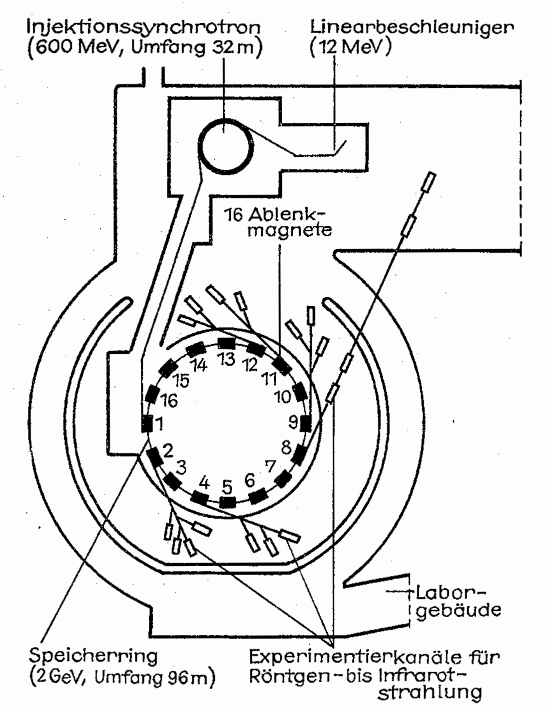


Abb. 7.1.6–2 „Lichtquelle" für Forschungszwecke: Elektronenspeicherring (Durchmesser 31 m) für 2-GeV-Elektronen (375 mA Maximalstrom) (Daresbury Lab., England)

günstiger als Synchrotrons sind Speicherringe, die z. B. von einem Booster-Synchrotron gespeist werden. Mit beispielsweise einem Durchmesser von 31 m, 2 GeV Energie, 375 mA Elektronenstrom und 10^{-8} Pa Druck sowie einem Spektralbereich von 0,02 nm bis zum fernen Infrarot (SRS Daresbury, Großbritannien, Abb. 7.1.6–2) gehören solche Apparaturen zu den größten gegenwärtig labormäßig betriebenen „Lichtquellen", die nur noch von einigen hauptsächlich für Kernforschungszwecke benutzten Anlagen besonders in der UdSSR und den USA übertroffen werden. Während Röntgen-Strahlen also auch an Synchrotrons genutzt werden können, werden sie in konventionellen Anlagen in abgeschmolzenen oder offenen Röntgen-Röhren erzeugt. Für Röntgen-Beugungsuntersuchungen sind gut fokussierte Elektronenstrahlen notwendig, um in den Feinfokusröhren eine geeignete Strahlung zu erhalten, wobei durch von außen magnetisch angetriebene „Drehanoden" eine besonders hohe Belastung und damit vergleichsweise große Strahlungsintensität ermöglicht wird. Für ESCA-Fotoelektronenspektrometer nutzt man die charakteristische Strahlung von Aluminium oder Magnesium aus, um (weitgehend) monochromatisches Röntgen-Licht zu erhalten. Auch im Röntgen-Gebiet gibt es Monochromatoren, wobei als „Beugungsgitter" ebene oder gebogene Einkristalle, z. B. LiF, SiO_2, Si, verwendet werden. Die Synchrotronröntgen-Strahlung wird außer für Fotoemission und Absorption besonders auch für Absorptionsfeinstrukturmessungen zur Bestimmung von Atomabständen benutzt (EXAFS). Im Gegensatz zu den spektroskopisch eingesetzten „weichen" Strahlen werden „harte" Röntgen-Strahlen wie auch γ-Strahlen zur Defektoskopie, z. B. Schweißnahtprüfung, verwendet. Hierfür sind Hochleistungsröntgen-Röhren bzw. γ-Strahlungspräparate erforderlich.
Die Erzeugung elektromagnetischer Strahlung im Bereich der Radiowellen ist von der Rundfunktechnik her bekannt und basiert auf abgestimmten Schwingkreisen und der Wellenlänge angepaßten Antennen. Besonderheiten treten hauptsächlich bei Mikrowellen auf, also im Bereich der dm-, cm-, mm- und Submillimeterwellen. Hier erfolgt die Schwingungserzeugung durch Spezialröhren oder Halbleiterbauelemente, wie Reflexklystrons, Wanderfeldröhren und -magnetrons, Gunn-Oszillatoren usw. Typisch ist die Hohlleitertechnik zur Fortleitung und Abstrahlung der elektromagnetischen Wellen, die z. B. für epr-Spektrometer erforderlich ist.

7.2. Metrologie

7.2.1. Aufgaben der Metrologie

Die Metrologie ist die *Lehre vom Messen* (vgl. 7.2.3.1.). Sieht man von methodischen Gesichtspunkten ab, sind in der Metrologie wie in allen Wissenschaften theoretische, praktische und organisatorische Aspekte zu unterscheiden. Daraus resultiert die Gliederung der Metrologie, wie sie Tab. 7.2.1–1 zeigt. Die Metrologie ist eine typische Querschnittswissenschaft, die alle Bereiche von Wissenschaft und Produktion bis hinein ins tägliche Leben durchdringt.
7.2.1.1. Einsatzbereiche der Meßtechnik. Dem Querschnittscharakter der Meßtechnik entsprechend sind ihre Einsatzgebiete vielfältig und kaum überschaubar. Faßt man diese – auch unter Beachtung der vorrangig angewendeten Methoden – sinnvoll zusammen, so ergibt sich das in Tab. 7.2.1–2 gezeigte Schema. Dabei sind die relativ spe-

ziellen Techniken von gut abgrenzbaren Einsatzgebieten unter dem Begriff *Spezialmeßtechnik* zusammengefaßt. Hierzu gehören die Meßtechnik für Astronomie, Geodäsie, Meteorologie, Medizin, Werkstoffkunde (Werkstoffprüfung), Militärwesen, Verkehr zu Lande, zu Wasser und in der Luft (einschließlich Raumfahrt), Sport u.a. Schon diese Aufzählung läßt erkennen, welche Bedeutung der Metrologie zukommt. Kaum eine Wissenschaft kann ohne eine gut entwickelte Meßtechnik auskommen, und die moderne Produktion einschließlich der Verteilung der Erzeugnisse (Handel) ist gleichfalls in hohem Maße auf die Meßtechnik angewiesen. Etwa 20 % des gesellschaftlichen Arbeitsvermögens, in der elektronischen Industrie sogar rund das Doppelte, werden in der Produktion für Meß- und Prüfprozesse zur Qualitätssicherung und Produktionssteuerung aufgewendet.
Oft wird in der Bezeichnung das Einsatzgebiet der jeweiligen Meßtechnik direkt zum Ausdruck gebracht, z.B. Fertigungs- oder Betriebsmeßtechnik. Dementsprechend gibt es auch eine physikalische Meßtechnik, die bestimmte Besonderheiten aufweist. Durch die Integration der Wissenschaften bedingt sind allerdings die Grenzen zu anderen Einsatzgebieten, wie zur chemischen Meßtechnik, fließend.

7.2.1.2. Physikalische Meßtechnik. Die physikalische Meßtechnik ist zunächst dadurch gekennzeichnet, daß die Wertebereiche der zu messenden Größen oft sehr beträchtlich sind. So reichen beispielsweise bei den Längen die Größenordnungen von den Werten, wie sie in der Kernphysik auftreten (z.B. Kerndurchmesser $\approx 10^{-15}$ m), bis zu astronomischen Werten (Lichtjahr $\approx 10^{15}$ m). Das sind etwa 30 Größenordnungen, die direkt oder indirekt meßtechnisch erfaßt werden müssen.
Nach dem Schema in Tab. 7.2.1–2 gehört die physikalische Meßtechnik sowohl zur Labormeßtechnik (Durchführung physikalischer Experimente), als auch zur Präzisionsmeßtechnik, z.B. Bestimmung von Naturkonstanten. Die durchschnittlichen Genauigkeitsforderungen sind demzufolge hoch bis sehr hoch. Dabei ist es vorteilhaft, daß physikalische Messungen normalerweise wiederholbar sind, weil die Meßbedingungen konstant gehalten oder reproduzierbar wiederhergestellt werden können, so daß durch die Aufnahme von Meßreihen der zufällige Fehler klein gehalten werden kann (vgl. 7.2.6.4.).
Daneben gibt es auf dem Gebiet der physikalischen Meßtechnik Aufgaben, wie sie eigentlich für die industrielle Meßtechnik typisch sind. Insbesondere bei der Kontrolle von Umweltbedingungen und Störeinflüssen (Temperatur, Luftdruck, Konstanz elektrischer Versorgungsspannungen, Störfelder usw.) werden kontinuierliche Meß- und Registrierverfahren (vgl. 7.2.7.1.) eingesetzt.

7.2.1.3. Gliederung der Meßverfahren. In einem Überblick über das Gesamtgebiet der Metrologie und die vielfältigen Einsatzgebiete der Meßtechnik muß auch auf die Möglichkeit der Gliederung nach den Meßgrößen, wie sie vor allem bei der Betrachtung der einzelnen Meßverfahren üblich ist, hingewiesen werden. Eine Gliederung nach den Größenarten entspricht auch weitgehend der für die Physik gebräuchlichen Unterteilung. Danach sind Meßverfahren für geometrische, mechanische, thermische, elektrische und magnetische Größen, für Größen der elektromagnetischen und korpuskularen Strahlung sowie für Stoffeigenschaften und Stoffzusammensetzungen zu unterscheiden. Eine solche Gliederung liegt auch in etwas modifizierter Form der Einteilung der folgenden Abschnitte (vgl. 7.3. bis 7.6.) zugrunde.
Nach der angewandten Methodik, speziell nach dem Charakter der Abbildungs- und Ausgangssignale, werden oft mehrere Gebiete unter speziellen Aspekten zusammengefaßt. Als Beispiele seien genannt: pneumatische Meßtechnik, elektrische und/oder elektronische Meßtechnik, elektrische (elektronische) Messung nichtelektrischer Größen (vgl. 7.4.3.) und optische Messung nichtoptischer Größen.

Tab. 7.2.1–1 Gliederungsschema der Metrologie

Bezeichnung	Aspekt	typische Arbeitsgebiete
Meßkunde (Meßtheorie)	theoretischer	Größen und Einheiten Theorien, Arbeitsmethoden Fehlerbetrachtung, Meßdynamik
Meßtechnik (Meßpraxis)	anwendungstechnisch	Meßmittel Meßverfahren, Durchführung und Auswertung von Messungen
Meßwesen (gesetzliche Metrologie)	juristisch-organisatorisch	gesetzliche Vorschriften Organisation und Planung Ökonomie

Tab. 7.2.1–2 Gliederungsschema der Meßtechnik

Bezeichnung	Aufgaben (Beispiele)
Präzisionsmeßtechnik*	Ermittlung von Naturkonstanten
Labormeßtechnik*	experimentelle Untersuchungen
industrielle Meßtechnik	Betriebs- und Fertigungsmeßtechnik
Spezialmeßtechnik	Meßtechnik in Astronomie, Geodäsie usw.
„Alltags"-Meßtechnik	Thermometer, Uhr, Waage

* typisch für physikalische Meßtechnik

7.2.2. Größen und Einheiten

7.2.2.1. Größen und Größenarten. Physikalische Größen. Eine physikalische Größe ist eine Eigenschaft eines Objektes bzw. einer Erscheinung, z.B. Gegenstand, Stoff, Vorgang, Zustand, die qualitativ charakterisiert und quantitativ beschrieben werden kann. Die qualitative Charakterisierung ist oft einfach und damit auch eindeutig, z.B. bei der Größe „Länge" (eindimensionale räumliche Ausdehnung), vielfach aber auch schwieriger, z.B. Energien in ihren verschiedenen Erscheinungsformen. Gleichheit von Dimensionen oder möglichen Einheiten ist kein Beweis für die Gleichartigkeit von Größen. Aus der Quantifizierbarkeit von Größen folgt die Möglichkeit, mit ihnen rechnen zu können, sie beinhaltet auch ihre prinzipielle Meßbarkeit. Schönheit und Behaglichkeit sind also keine Größen, obwohl es sich um charakterisierbare Eigenschaften handelt.
Formelzeichen. Da es unzweckmäßig wäre, die Namen der physikalischen Größen z.B. in Formeln auszuschreiben, hat man sich auf bestimmte Formelzeichen (Symbole) geeinigt.

Das sind kleine und große Buchstaben, die im Druck kursiv gesetzt werden, um sie von anderen Abkürzungen, Kurzzeichen und Symbolen zu unterscheiden. Es bedeuten also *m* eine Masse, *F* eine Kraft usw. Da es mehr physikalische Größen als Buchstaben gibt, ist es unvermeidlich, daß verschiedentlich mehrere Größen das gleiche Formelzeichen haben. Kann dies zu Verwechslungen führen, sind Schriften mit verschiedenem Schriftcharakter oder Schriftschnitt zu benutzen oder von Groß- auf Kleinbuchstaben bzw. umgekehrt auszuweichen. Durch Indizes sollten nur verschiedene Größen gleicher Größenart unterschieden werden. Vektorielle Größen werden durch Fettdruck, darüber gesetzten Pfeil oder Frakturbuchstaben gekennzeichnet.

Größensysteme. Physikalische Größen stehen untereinander im Zusammenhang. Um derartige Zusammenhänge formelmäßig beschreiben zu können, müssen einige Größen als Basisgrößen festgelegt und verbal erklärt werden, soweit sie nicht durch Erfahrung als bekannt vorausgesetzt werden können. Alle anderen Größen sind dann durch Formeln aus ihnen ableitbar (*abgeleitete Größen*). Mit den Basisgrößen Länge l und Zeit t wird die abgeleitete Größe *gleichförmige Geschwindigkeit* v durch die Gleichung $v = l/t$ definiert. Wählt man die Anzahl der Basisgrößen eines Größensystems zu klein, so lassen sich mehrere abgeleitete Größen, die nach unserer Erfahrung qualitativ verschieden sind, durch die gleiche Formel definieren. Im umgekehrten Fall (zu viele Basisgrößen) tauchen in den Definitionsgleichungen der abgeleiteten Größen zusätzliche dimensionsbehaftete (physikalisch nicht ohne weiteres interpretierbare) Konstanten auf, die das Verständnis und die Übersicht erschweren. Die Mechanik läßt sich beispielsweise mit drei Basisgrößen widerspruchsfrei und unseren Erfahrungen entsprechend beschreiben. Die Gesamtheit der Basisgrößen und der aus ihnen abgeleiteten Größen bilden ein *Größensystem*.

Größenarten und Dimensionen. Die Gesamtheit aller Größen gleicher Art bildet eine *Größenart*. So gehören Abstände, Dicken, Durchmesser, Höhen usw. zur Größenart Länge. Größen gleicher Größenart lassen sich (zumindest formal) addieren, was bei unterschiedlicher Größenart nicht möglich ist. (Zeiten lassen sich nicht zu Temperaturen addieren). Leistung als Arbeit/Zeit ist also eine Größenart, die Leistung von Elektromotoren ist eine *allgemeine Größe*, und die Leistung eines speziellen Motors, deren konkreter Wert auch gemessen werden kann, ist eine *spezielle Größe*. Da die wertmäßige Angebbarkeit das Charakteristikum spezieller Größen ist, sind auch Mittelwerte, Nennwerte usw. spezielle Größen.

Von den Größenarten zu unterscheiden sind die *Dimensionen*. Sie sind die Potenzprodukte aus Basisgrößen, die sich aus der Definitionsgleichung einer abgeleiteten Größe ergeben, und werden $\dim G$ (= Dimension der Größe G) abgekürzt. Die Dimensionen der Basisgrößen erscheinen im Druck grotesk. Sind also L und T die Dimensionen von Länge und Zeit, so wird die Dimension der gleichförmigen Geschwindigkeit durch die Gleichung $\dim v = \mathsf{LT}^{-1}$ angegeben. Es kommt vor, daß auch im gleichen Größensystem unterschiedliche Größenarten die gleiche Dimension haben. Beispielsweise ist die Dimension einer Arbeit W (aus der Kraft F längs des Weges l) und des Drehmomentes M, wenn eine Kraft F senkrecht zum Radius r wirkt, gleich: $\dim W = \dim Fl = \mathsf{ML^2T^{-2}} = \dim Fr = \dim M$.

Deshalb ist es eindeutiger, wenn der Begriff Größenart verwendet wird. Eine besondere Rolle spielen die *Größen der Dimension 1*, die auch oft *dimensionslose Größen* genannt werden. Dazu gehören Zählgrößen, Kennzahlen (z. B. Reynolds-Zahl, vgl. 2.2.5.2.), Quotienten aus Größen gleicher Größenart (z. B. Quotient aus der Lichtgeschwindigkeit im Vakuum und in einem Medium = [absolute] Brechzahl, vgl. 2.5.2.1.) und Logarithmen von Verhältnisgrößen (z. B. Dämpfungen, Pegel, vgl. 7.2.2.4.).

7.2.2.2. Einheiten und Einheitensysteme. Definition der Einheit. Um den Wert einer konkreten Größe zahlenmäßig angeben zu können, benötigt man eine Vergleichsgröße mit einem festgelegten Wert. Eine solche Vergleichsgröße, deren Wert jetzt allgemein international vereinbart ist, heißt *Einheit*, (wenn Verwechslungen möglich sind, auch *Maßeinheit*). Die Einheit einer Größe wird allgemein angegeben, indem man das Formelzeichen in eckige Klammern setzt; die Einheit der Masse m ist allgemein durch [m] auszudrükken. Der Quotient aus dem Wert einer konkreten Größe und ihrer Einheit heißt *Zahlenwert* (früher auch *Maßzahl*) und wird in allgemeiner Schreibweise durch das Formelzeichen der Größe in geschweiften Klammern ausgedrückt. Jede konkrete Größe hat einen bestimmten Wert, der also, wenn er bekannt ist, als Produkt aus Zahlenwert und Einheit angegeben werden kann. Daraus resultiert ein weiteres Kriterium für Größen: Alles, was sich durch ein Produkt *Zahlenwert* × *Einheit* ausdrücken läßt, ist eine (spezielle) Größe. Die Dicke d_{St} eines Stabes ist also gegeben durch $d_{St} = \{d_{St}\}\,[d_{St}]$. Die Aufgabe einer Messung besteht somit darin, $\{G\}$ zu ermitteln, nachdem für G eine konkrete Einheit vereinbart worden ist. Ähnlich wie für Größen sind auch für Einheiten Kurzzeichen vereinbart, die *Einheitenzeichen*, die im Druck im Gegensatz zu den Formelzeichen gerade stehen. Dadurch werden Verwechslungen vermieden, denn es ist eindeutig, daß mit *m* eine Masse, mit m dagegen die Einheit Meter gemeint ist.

Tab. 7.2.2-1 Basiseinheiten des SI

Größe	Einheitenname	Einheitenzeichen	Grundlage der Definition
Länge	Meter	m	Wellenlänge der Strahlung von ^{86}Kr; laut Beschluß der 17. Generalkonferenz für Maß und Gewicht vom 20.10.1983: Lichtgeschwindigkeit (vom Licht zurückgelegter Weg)
Masse	Kilogramm	kg	internationaler Prototyp
Zeit	Sekunde	s	Periodendauer der Strahlung von ^{133}Cs
elektrische Stromstärke	Ampere	A	Anziehungskraft einer stromdurchflossenen Doppelleitung
(thermodynamische) Temperatur	Kelvin	K	Tripelpunkt des Wassers
Lichtstärke	Candela	cd	Strahlung der Frequenz $540 \cdot 10^{12}$ Hz
Stoffmenge	Mol	mol	Anzahl von Atomen in 12 g Kohlenstoff^{12}

Einheitensysteme. Ebenso wie die Größen sind auch die Einheiten nicht voneinander unabhängig, sondern die meisten Einheiten sind durch Definitionsgleichungen aus einigen verbal definierten *Basiseinheiten* abzuleiten. Manche dieser abgeleiteten Einheiten haben einen eigenen Namen (vgl. Tab. 7.2.2–2.). Die Gesamtheit der Basiseinheiten und abgeleiteten Einheiten bildet ein *Einheitensystem* (früher *Maßsystem* genannt).

Im Laufe der Zeit gab es schon viele Einheitensysteme. Bekannt und teilweise noch verbreitet ist das *physikalische Maßsystem*, das nach den verwendeten Basiseinheiten der Mechanik Zentimeter (cm), Gramm (g) und Sekunde (s) auch *c-g-s-System* heißt. Besonders in der Elektrizitätslehre wurden früher sehr viele unterschiedliche Einheitensysteme benutzt; zeitweilig gab es für die elektrische Spannung und für den elektrischen Strom verschiedene Einheiten (häufig sogar mit gleichem Namen aber unterschiedlichem Wert).

Das jetzt international überwiegend benutzte Einheitensystem ist das *Système International d'Unités*, das in allen Sprachen SI (nicht SI-System) abgekürzt wird. Der große Vorteil des SI ist neben seiner Internationalität seine Kohärenz, d. h., es besteht nur aus Basiseinheiten und kohärenten Einheiten. *Kohärente Einheiten* sind solche, die als Potenzpunkt aus den Basiseinheiten abgeleitet sind, wenn dabei nur der Zahlenfaktor 1 auftritt.

7.2.2.3. Das Internationale Einheitensystem SI. Basiseinheiten. Die sieben Basiseinheiten des SI sind in Tab. 7.2.2–1 zusammengestellt. Nur eine von ihnen (das Kilogramm) wird durch einen *Prototyp* verkörpert, alle anderen gehen auf Naturkonstanten zurück. Für die Temperatureinheit Kelvin gibt es noch eine besondere Benennung, das Grad Celsius (°C) als Einheit der Differenz aus einer (absoluten) Temperatur T und der Temperatur $T_C = 273{,}15$ K. Die *Celsius-Temperatur* t oder φ ist also $t = T - T_C$. Für Temperaturdifferenzen ist auch bei Celsius-Temperaturen Kelvin als Einheit zu bevorzugen. Die Meßspanne (vgl. 7.2.5.1.) eines Thermometers, dessen Meßbereich von −40 °C bis +40 °C reicht, beträgt also 80 K. Da die Basiseinheit der Masse bereits einen Vorsatz hat, das *Kilo*gramm, werden dezimale Vielfache und Teile durch Vorsätze beim Gramm (= 10^{-3} kg) gebildet.

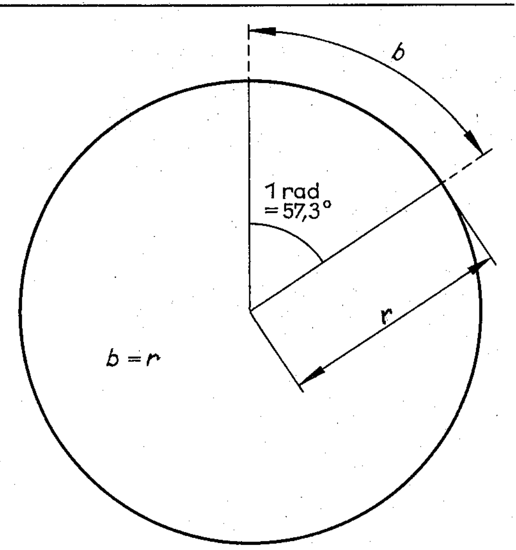


Abb. 7.2.2–3 Geometrische Darstellung der SI-Einheit Radiant

Abgeleitete Einheiten werden mit Hilfe der Basiseinheiten durch *Einheitengleichungen* definiert. Die Einheit der Kraft F ist entsprechend der Beziehung $F = ma$ (Beschleunigung $a = l t^{-2}$) durch die Einheitengleichung $[F] = \text{kg} \cdot \text{m} \cdot \text{s}^{-2}$ definiert. Sie hat den Namen Newton (N), also $[F] = \text{N}$. Es gibt noch weitere abgeleitete Einheiten mit eigenen Namen. Die meisten wurden nach bekannten Naturwissenschaftlern oder Technikern benannt (Tab. 7.2.2–2).

Die SI-Einheit des ebenen Winkels ist der Radiant (rad). Es ist dies der ebene Winkel zwischen zwei Kreisradien, die aus dem Umfang einen Bogen b von der Länge des Radius r ausschneiden (Abb. 7.2.2–3). Nach dieser verbalen Definition ist Radiant als Basiseinheit anzusehen. Da aber Bogenlänge und Radius in gleichen Einheiten $[b] = \text{m}$ und $[r] = \text{m}$

Tab. 7.2.2–2 Abgeleitete SI-Einheiten mit eigenen Namen

Größe	Einheitennamen	Einheitenzeichen	Definition	Benennung nach
Kraft	Newton	N	$1\,\text{N} = 1\,\text{m} \cdot \text{kg} \cdot \text{s}^{-2}$	Isaac Newton
Druck	Pascal	Pa	$1\,\text{Pa} = 1\,\text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-2}$	Blaise Pascal
Arbeit	Joule	J	$1\,\text{J} = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2}$	James P. Joule
Leistung	Watt	W	$1\,\text{W} = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3}$	James Watt
Frequenz	Hertz	Hz	$1\,\text{Hz} = 1\,\text{s}^{-1}$	Heinrich Hertz
elektrische Ladung	Coulomb	C	$1\,\text{C} = 1\,\text{s} \cdot \text{A}$	Charles A. de Coulomb
elektrische Spannung	Volt	V	$1\,\text{V} = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-1}$	Alessandro Volta
elektrische Kapazität	Farad	F	$1\,\text{F} = 1\,\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^4 \cdot \text{A}^2$	Michael Faraday
elektrischer Widerstand	Ohm	Ω	$1\,\Omega = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-3} \cdot \text{A}^{-2}$	Georg S. Ohm
elektrischer Leitwert	Siemens	S	$1\,\text{S} = 1\,\text{m}^{-2} \cdot \text{kg}^{-1} \cdot \text{s}^3 \cdot \text{A}^2$	Werner v. Siemens
magnetischer Fluß	Weber	Wb	$1\,\text{Wb} = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$	Wilhelm E. Weber
magnetische Induktion	Tesla	T	$1\,\text{T} = 1\,\text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-1}$	Nikola Tesla
Induktivität	Henry	H	$1\,\text{H} = 1\,\text{m}^2 \cdot \text{kg} \cdot \text{s}^{-2} \cdot \text{A}^{-2}$	Joseph Henry
Energiedosis	Gray	Gy	$1\,\text{Gy} = 1\,\text{m}^2 \cdot \text{s}^{-2}$	Louis H. Gray
Aktivität	Becquerel	Bq	$1\,\text{Bq} = 1\,\text{s}^{-1}$	Henri Becquerel
Lichtstrom	Lumen	lm	$1\,\text{lm} = 1\,\text{cd} \cdot \text{sr}$	(lat., Licht)
Beleuchtungsstärke	Lux	lx	$1\,\text{lx} = 1\,\text{m}^{-2} \cdot \text{cd} \cdot \text{sr}$	(lat., Tageslicht)

gemessen werden, also von gleicher Größenart sind (wobei der Unterschied zwischen Kreisbogen und Gerade vernachlässigt ist), kann man den Radiant gemäß der Gleichung 1 rad = 1 m/1 m = 1 auch als abgeleitete Einheit auffassen. Ganz ähnlich ist es mit der Einheit des Raumwinkels Ω, dem Steradiant (sr), 1 sr = 1 m²/1 m². Ω ist der Raumwinkel, der von einer Radienschar gebildet wird, die auf der Oberfläche einer Kugel eine Fläche herausschneidet, die gleich r^2 ist. rad und sr bezeichnet man deshalb als *ergänzende Einheiten*, neigt aber jetzt dazu, sie den abgeleiteten Einheiten zuzurechnen.

Dezimale, Vielfache und Teile von Einheiten. Um bei sehr kleinen bzw. sehr großen Werten konkreter Größen nicht zu unbequem großen Zahlenwerten zu kommen, werden (negative und positive) Zehnerpotenzen der Einheiten durch *Vorsätze* gebildet. Tab. 7.2.2-4 enthält alle Vorsätze für die Bildung dezimaler Vielfacher und Teile von Einheiten. Die Vorsätze Hekto, Deka, Dezi und Zenti sollen nur noch dort verwendet werden, wo es bisher schon üblich war; zulässig ist z. B. Zentimeter (cm), aber nicht Zentiampere (cA).

Tab. 7.2.2-4 Vorsätze für Einheiten (der Faktor gibt an, um wieviel der Wert der Einheit vergrößert wird)

Vorsatz	Vorsatzzeichen	Faktor
Exa	E	10^{18}
Peta	P	10^{15}
Tera	T	10^{12}
Giga	G	10^{9}
Mega	M	10^{6}
Kilo	k	10^{3}
Hekto	h	10^{2}
Deka	da	10
Dezi	d	10^{-1}
Zenti	c	10^{-2}
Milli	m	10^{-3}
Mikro	µ	10^{-6}
Nano	n	10^{-9}
Piko	p	10^{-12}
Femto	f	10^{-15}
Atto	a	10^{-18}

Vorsatzzeichen und Einheitenzeichen werden zusammengeschrieben und auch gemeinsam zur Potenz erhoben.

Einige Vielfache und Teile von SI-Einheiten haben – historisch bedingt – eigene Namen, die, obwohl sie inkohärent sind, weiterbenutzt werden dürfen. Es sind dies

Liter (l) für 1 dm³ = $1 \cdot 10^{-3}$ m³,

Tonne (t) für $1 \cdot 10^{3}$ kg,

Bar (bar) für $1 \cdot 10^{5}$ Pa.

Das Bar wird ungültig, sobald entsprechende internationale Beschlüsse vorliegen.

7.2.2.4. Gültige Einheiten. Mit den SI-Einheiten und ihren Vielfachen und Teilen allein kommt man nicht aus, insbesondere weil fest eingewurzelte Gewohnheiten dagegen stehen. Deshalb sind außerhalb des SI noch einige weitere *gültige Einheiten* zugelassen. Neben den in 7.2.2.3. genannten sind dies die zeitlich und hinsichtlich des Anwendungsbereiches nicht eingeschränkten Zeiteinheiten Minute (min), Stunde (h) und Tag (d), die Kalendereinheiten Woche, Monat und Jahr (a) sowie die Einheiten des ebenen Winkels Grad (°), Minute (′) und Sekunde (″).

In bestimmten Anwendungsgebieten sind einige weitere *SI-fremde Einheiten* unbefristet gültig. Für die Physik von Interesse sind u. a. in der Atom- und Kernphysik die Atomare Masseneinheit u (1 u = $1{,}660\,57 \cdot 10^{-27}$ kg) und die Energieeinheit Elektronenvolt eV (1 eV = $1{,}602\,19 \cdot 10^{-19}$ J), in der Astronomie die Astronomische Einheit AE (1 AE = $1{,}495\,98 \cdot 10^{11}$ m) und das Lichtjahr ly (1 ly = $0{,}946\,05 \cdot 10^{16}$ m) sowie in der Optik die Dioptrie dpt (1 dpt = 1 m⁻¹). Außer beim Elektronenvolt sind bei allen in diesem Abschnitt genannten Einheiten, ebenso wie bei °C, keine Vorsätze zugelassen.

Weitere gültige Einheiten sind noch jene für *Verhältnisgrößen* Prozent (%), Promille (‰), Millionstel (ppm) und Milliardstel (ppb), die besonders für Konzentrationsangaben benötigt werden. Für die logarithmierten Verhältnisgrößen sind die Kennwörter Bel (B) bei dekadischen und Neper (Np) bei natürlichen Logarithmen zugelassen. Durch Logarithmieren zur Basis 10 ergibt sich aus dem Verhältnis zweier Leistungsgrößen P (Größen mit quadratischer Abhängigkeit) bzw. dem Verhältnis der Quadrate zweier Feldgrößen A (Größen mit linearer Abhängigkeit, z. B. Spannungen) der Wert 1 Bel. In der Praxis wird vorzugsweise Dezibel (1 dB = 0,1 B) benutzt, wofür gilt

$$10\left(\lg\frac{P_1}{P_2}\right)\text{dB} = 10\left(\lg\frac{A_1^2}{A_2}\right)\text{dB} = 20\left(\lg\frac{A_1}{A_2}\right)\text{dB}.$$

Tab. 7.2.2-5 verdeutlicht die Anwendung des Dezibel.

7.2.2.5. Maßverkörperungen. Darstellung von Einheiten. Abgesehen vom Kilogramm (vgl. 7.2.2.3.) sind alle Basiseinheiten des SI durch Naturkonstante bzw. Meßvorschriften definiert. Das bedeutet, die Definition der jeweiligen Einheit ist in einem speziellen Laboratorium durch eine geeignete Meßeinrichtung mit hoher Genauigkeit zu realisieren, um sie auf ein geeignetes Meßgerät oder eine Maßverkörperung übertragen zu können; man spricht von einer (fundamentalen) *Darstellung der Einheit*. Eine *Maßverkörperung* ist ein Meßmittel, das einen oder auch mehrere Werte einer Größe verkörpert. Die aufwendigen Meßeinrichtungen zur Darstellung der Einheiten heißen *Primärnormale*. Sie werden nur in größeren Abständen an wenigen Stellen (meist in metrologischen Staatsinstituten) betrieben.

Der absolute Wert einer konkreten Meßgröße kann nie mit

Tab. 7.2.2-5 Auswahl logarithmierter Verhältnisgrößen (dekadischer Logarithmen)

Dezibel	Verhältnis zweier Leistungsgrößen	Feldgrößen
0,0	1,000	1,000
0,1	1,023	1,012
0,2	1,047	1,023
0,5	1,122	1,059
1,0	1,259	1,122
2,0	1,585	1,259
3,0	1,995	1,413
5,0	3,162	1,778
10,0	10,00	3,162
20,0	10^{2}	10,00
40,0	10^{4}	10^{2}
100,0	10^{10}	10^{5}

Tab. 7.2.2–6 Auswahl erreichbarer Meßunsicherheiten

Einheit	relative Meßunsicherheit bei Darstellung	der Einheit von Sekundärnormalen
Meter	10^{-9} bis 10^{-8}	$5 \cdot 10^{-8}$
Kilogramm	(Definition!)	$3 \cdot 10^{-8}$
Sekunde	$5 \cdot 10^{-13}$	$2 \cdot 10^{-12}$
Ampere	(2 bis 4) $\cdot 10^{-6}$	$5 \cdot 10^{-6}$
Kelvin	abhängig von T	0 bis 900 K: 0,2 K und besser
Candela	$2 \cdot 10^{-3}$	etwa 10^{-2}

einer kleineren Meßunsicherheit ermittelt werden, als die Darstellung der benutzten Einheit möglich ist. Deshalb muß das Bestreben dahin gehen, bei der Darstellung der Einheiten die höchstmögliche Genauigkeit zu erreichen. Tab. 7.2.2–6 gibt einige orientierende Werte für die derzeit erreichbaren Meßunsicherheiten bei der Darstellung der Einheiten. Aus den Angaben in dieser Tabelle wird z. B. verständlich, warum bei hochgenauen Messungen das Bestreben dahin geht, die zu messenden Größen auf Zeitintervalle oder Frequenzen abzubilden, d. h., primär in derartige Größen umzuwandeln.

Bewahrung und Weitergabe von Einheiten. Mit Hilfe der Primärnormale werden die Gerätekonstanten von zeitlich hochkonstanten Meßmitteln bestimmt, die zur Bewahrung der Einheiten dienen. Diese heißen *Sekundärnormale*. An die Primär- und Sekundärnormale werden weitere Normale angeschlossen, die *Referenznormale* 1., 2. Ordnung usw. Diese Referenznormale sind es, mit deren Hilfe die Arbeitsmeßmittel auf Richtigkeit geprüft werden können. Das Anschließen niederer an höhere Normale erfolgt durch einen Vergleich, bei dem natürlich Genauigkeit eingebüßt wird. Die Meßunsicherheit der Referenznormale liegt also deutlich über derjenigen der Primärnormale. Tab. 7.2.2–7 zeigt als Beispiel die Fehlerklassen von Drucknormalen (Kolbenmanometern). Dafür ist aber auch der Aufwand für die Meßverfahren bei den niedrigeren Normalen geringer.

Die Sekundär- und Referenznormale zur Bewahrung und Weitergabe von Einheiten unterscheiden sich vielfach in ihrem Meßverfahren bzw. in der Art der Einheitenverkörperung von den Primärnormalen.

Zur Bewahrung der Stromstärkeeinheit dienen in erster Linie Maßverkörperungen für Spannung und Widerstand. Als Spannungsnormale verwendet man *Normalelemente*, und zwar Internationale Weston-Elemente. Ihre EMK liegt zwischen 1,018 55 und 1,018 65 V und ändert sich höchstens um etwa 1 µV/a. Als *Normalwiderstände* dienen Drahtspulen aus Manganin (CuMnNi-Legierung), die in Büchsen eingeschlossen sind. Sie ändern ihren Wert um weniger als 1 µΩ/a, und ihr Temperaturkoeffizient beträgt weniger als $1 \cdot 10^{-5}\,K^{-1}$ bei 20 °C.

Zur Bewahrung der Temperatureinheit Kelvin wird eine sog. „konventionelle Skale", die empirische *Internationale praktische Temperaturskale* (IPT-Skale), benutzt. Sie stützt sich auf sechs primäre (fundamentale) und mehrere sekundäre Fixpunkte natürlicher Gleichgewichtstemperaturen (meist Erstarrungs- und Siedepunkte) und auf Interpolationsgleichungen zwischen diesen Punkten. Ein fundamentaler Fixpunkt ist der Tripelpunkt des Wassers bei 273,16 K (0,01 °C).

Tab. 7.2.2–7 Fehlerklassen von Drucknormalen (Kolbenmanometer)

Normal	relative Meßunsicherheit
Primärnormal	$2{,}5 \cdot 10^{-5}$
Sekundärnormal	(3 bis 8) $\cdot 10^{-5}$
Referenznormal erster Ordnung	$\leqq 2 \cdot 10^{-4}$
Referenznormal zweiter Ordnung	$\leqq 5 \cdot 10^{-4}$
Referenznormal dritter Ordnung	$\leqq$ (1 bis 2) $\cdot 10^{-3}$
Referenznormal vierter Ordnung*	$\leqq$ (4 bis 6) $\cdot 10^{-3}$
Arbeitsmeßmittel*	$\leqq$ (0,5 bis 6) $\cdot 10^{-2}$

* Federmanometer

7.2.3. Metrologische Tätigkeiten

7.2.3.1. Messen und Prüfen. Messen heißt, den unbekannten Wert einer physikalischen Größe auf experimentellem Wege zu ermitteln. Dies geschieht stets durch einen *Vergleich* mit der Einheit der betreffenden Größenart. Nur wenn eine Einheit definiert und für den konkreten Fall festgelegt ist, hat der Zahlenwert, der durch die Messung ermittelt wird, einen Sinn. Die Einheit muß direkt oder indirekt in irgendeiner Form in der benutzten Meßanordnung enthalten (verkörpert) sein. Direkte Maßverkörperungen der Einheiten finden sich nur in wenigen Meßanordnungen (z. B. die Wägestücke einer Hebelwaage, vgl. 7.3.2.); normalerweise wird die Einheit indirekt verkörpert, und zwar durch eine Größe, die mit der Einheit in gesetzmäßigem Zusammenhang steht.

Das Vergleichen ist nach dem Obengesagten notwendiger Bestandteil jedes Meßvorgangs. Nach der Vorgehensweise unterscheidet man verschiedene Meßmethoden.

Meßmethoden. Während mit *Meßprinzip* die physikalische Erscheinung bezeichnet wird, die einer Messung zugrunde liegt (z. B. wird beim Thermoelement der thermoelektrische Effekt als Meßprinzip benutzt), nennt man die Art und Weise, in der bei einer Messung der Vergleich realisiert wird, *Meßmethode*. Sie ist unabhängig vom angewandten Meßprinzip und von der zu messenden Größe. Es gibt viele verschiedene Meßmethoden, und ihre Anzahl nimmt mit der technischen Entwicklung ständig zu. Deshalb werden hier nur einige Beispiele für verbreitete Meßmethoden angegeben, um den Begriff weiter zu verdeutlichen: *Differenzmeßmethode* (Differenzmethode). Statt des absoluten Wertes der Meßgröße wird die Differenz zwischen dem Wert der Meßgröße und einem bekannten, konstanten Wert gleicher Größenart gemessen.

Nullabgleich-Meßmethode. Differenzmethode, bei der durch definierte Veränderung des Vergleichswertes die Differenz zwischen den beiden Werten beseitigt wird.

Kompensationsmeßmethode (Kompensationsmethode). Nullmethode bei energietragenden Größen, bei der aus der Einstellung der Vergleichsgröße (hier *Kompensationsgröße*) auf den Wert der Meßgröße geschlossen wird.

Substitutionsmeßmethode. Meßmethode, bei der nacheinander die Meßgröße und eine Maßverkörperung der Größe mit bekanntem Wert an den Eingang der Meßeinrichtung, den *Meßfühler*, angeschlossen wird.

Die Anwendung eines Meßprinzips und einer Meßmethode für die Messung einer konkreten Meßgröße bezeichnet man als *Meßverfahren*. Es ist stets meßgrößenbezogen (z. B. Temperaturmeßverfahren mit ...).

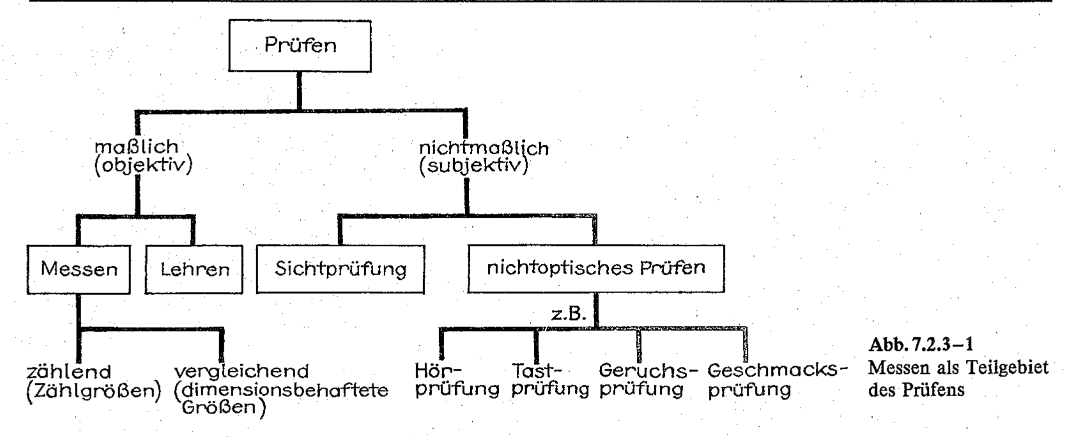


Abb. 7.2.3–1 Messen als Teilgebiet des Prüfens

Prüfen. Der Begriff „Prüfen" ist umfassender als „Messen". Er bezeichnet jeden Vorgang, durch den ermittelt wird, ob eine bestimmte Eigenschaft eines Prüflings, ein *Prüfparameter* oder *Prüfmerkmal*, vorgegebenen Anforderungen genügt. Auch das Prüfen beinhaltet also einen Vergleich, der aber nicht unbedingt objektiver Natur zu sein braucht. Deshalb unterscheidet man *maßliches* und *nichtmaßliches* (früher objektives und subjektives) *Prüfen* (Abb. 7.2.3–1). Die Prüfparameter des maßlichen Prüfens heißen *Variablenmerkmale*, die des nichtmaßlichen Prüfens *attributive Merkmale*. Das maßliche Prüfen ist entweder Messen bzw. Zählen oder „Lehren". Beim Lehren wird festgestellt, ob der Wert des Prüfparameters zwischen zwei vorgegebenen Grenzen liegt oder eine Grenze nicht über- bzw. unterschreitet. Diese Feststellung wird durch einen Vergleich des Prüflings mit einer Lehre getroffen, die ebenfalls eine Maßverkörperung darstellt. Lehren werden vor allem in der mechanischen Fertigung zur Prüfung geometrischer Parameter eingesetzt.

7.2.3.2. Kalibrieren, Graduieren und Justieren. Kalibrieren heißt, die Anzeigen bzw. sonstigen Ausgabewerte eines Meßmittels den einzelnen Werten der Meßgröße zuzuordnen. Ein neues Thermoelement wird z. B. kalibriert, indem festgestellt wird, welche Thermospannung bei welcher Temperatur angezeigt wird. Man nennt diesen Vorgang auch *Einmessen* oder Ermittlung der statischen Kennlinie (vgl. 7.2.5.1.). Auch die Aufnahme der Fehlerkennlinie eines Meßgerätes heißt kalibrieren. Sie gestattet es, für jeden abgelesenen Wert durch eine Korrektion den tatsächlichen Wert der Meßgröße (natürlich im Rahmen der Meßunsicherheit) zu bestimmen.

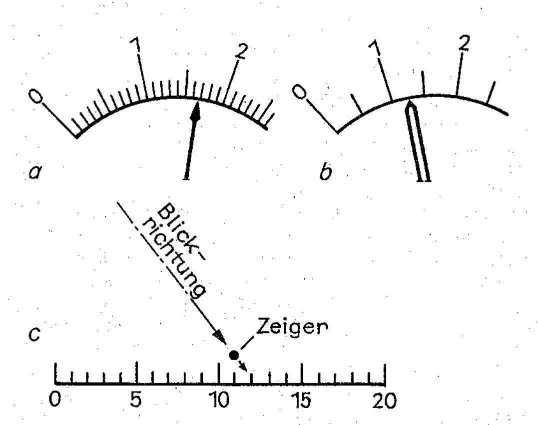


Abb. 7.2.3–2 Beispiele für Ablesefehler; (a) grober Ablesefehler: 1. Fehlablesung, z. B. 1,8 Skt, 2. Fehlinterpretation, z. B. bei Meßbereich 0 bis 45 mA; falsche Skalenkonstante 0,667 statt 1,5 mA · Skt^{-1} ergibt 11,33 mA statt 25,5 mA; (b) Ableseunsicherheit: x_m zwischen 1,2 und 1,3 Skt; (c) Parallaxenfehler (nicht senkrechte Blickrichtung): x_m 12 statt 11 Skt

Graduieren ist die experimentelle Bestimmung und das Auftragen der Teilungsmarken einer Meßmittelskale. Wird also im Beispiel des Thermoelementes beim Einmessen die Thermospannung durch ein Spannungsmeßgerät ohne Skaleneinteilung gemessen und bei jedem Ausschlag (z. B. von Grad Celsius zu Grad Celsius) ein Skalenstrich angebracht und die Temperatur dazugeschrieben, so ist dies ein Graduieren eines elektrischen Thermometers.

Justieren bedeutet, ein Meßmittel in einen solchen Zustand zu versetzen, daß es entsprechend den gegebenen Vorschriften meßbereit ist und seine Fehler innerhalb der angegebenen Grenzen liegen. Justieren einer Hebelwaage heißt also, sie verwackelungs- und erschütterungsfrei aufzustellen, sie mit einer Libelle senkrecht auszurichten und die Hebellängen zu korrigieren. Man nennt das Justieren auch *Abgleichen*, insbesondere dann, wenn die richtige Einstellung des Gerätes durch die Kontrolle seiner Anzeige bei einem oder zwei genau bekannten Werten der Eingangsgröße vorgenommen wird.

7.2.3.3. Ablesen und Auswerten. Ablesen. Der Abschluß des eigentlichen Meßvorganges ist die Ausgabe des unkorrigierten (oder korrigierten, vgl. 7.2.8.) Meßwertes (vgl. 7.2.7.), der bei analoger Anzeige von einer Skale abgelesen werden kann, falls dieser Meßwert nicht automatisch weiterverarbeitet wird. Dabei kommen auch *Ablesefehler* (Abb. 7.2.3–2) vor. Können nur Skalenteile (Skt) abgelesen werden, so müssen diese mit der Skalenkonstanten multipliziert werden, um den Meßwert zu erhalten. Die *Skalenkonstante* (SK) ergibt sich als Quotient aus Meßbereichsendwert und Wert des letzten Skalenteiles. Hat z. B. die Skale eines Strommeßgerätes 100 Skt, und der Meßbereich geht von 0 bis 5 A, dann ist SK = 5 A/100 Skt = 0,05 A/Skt. Liest man 40 Skt ab, so beträgt der Meßwert 40 · 0,05 A = 2 A.

Auswerten. Normalerweise ist eine Messung mit dem Able-

sen des Meßwertes noch nicht abgeschlossen; es muß sich eine *Auswertung* anschließen, um zum endgültigen Meßergebnis (vgl. 7.2.6.5.) zu gelangen. Zum Auswerten gehören insbesondere
- eine *Plausibilitätsbetrachtung*, ob der abgelesene Wert prinzipiell stimmen könnte (Abschätzen der möglichen Größenordnung, damit grobe Fehler ausgeschlossen werden, vgl. 7.2.6.2.),
- das Anbringen von *Korrektionen* für systematische Fehler und bekannte Einflußgrößen, z.B. Temperaturkorrektur,
- das Bestimmen der Meßunsicherheit bei Einzelmessungen, wenn die Messung nicht wiederholt werden kann oder soll,
- die Errechnung von Mittelwerten bei mehreren Messungen unter Wiederholbedingungen (Meßreihen),
- die Errechnung von Größen, die funktional von mehreren verschiedenen Meßgrößen abhängen, z.B. Dichte aus gemessener Masse und Volumen,
- die Bestimmung der Meßunsicherheit für Mittelwerte bzw. abhängige Größen (Fehlerfortpflanzung, vgl. 7.2.6.5.),
- die Formulierung des vollständigen Meßergebnisses unter Angabe der statistischen Sicherheit (vgl. 7.2.6.).

Mit dem Vorliegen des vollständigen Meßergebnisses ist die gesamte Messung beendet, und das Ergebnis kann genutzt werden (Vergleich mit theoretischen Werten, Klassieren, Ableiten weiterer Meßaufgaben usw.).

7.2.3.4. Eichen. Eichen ist eine metrologische Tätigkeit, die infolge ihres amtlichen Charakters eine Sonderstellung einnimmt. Man versteht darunter das von einer dafür eingesetzten Behörde durchgeführte Prüfen eines eichfähigen Meßmittels, das für Zwecke verwendet wird, für die geeichte Meßmittel vorgeschrieben sind. Dabei wird festgestellt und beurkundet, daß das Meßmittel den Anforderungen besonderer Vorschriften, den *Eichvorschriften*, genügt. Der *Eichpflicht* unterliegen Meßmittel, die eingesetzt werden
- im rechtsgeschäftlichen Verkehr zur Quantitäts- und Qualitätsbestimmung,
- im Gesundheitsschutz, Arbeitsschutz, Sicherheitswesen sowie bei der Überwachung des Straßenverkehrs,
- bei Erstattung von Gutachten und der Durchführung von Musterprüfungen.

Die Eichung ist nur für eine bestimmte Zeit (oft 2 Jahre) gültig, danach muß das Meßmittel nachgeeicht werden. Auch nach Beschädigungen und Reparaturen ist eine *Nacheichung* erforderlich. Die amtliche Prüfung und Beurkundung der ordnungsgemäßen Funktion von *Normalen* ist ebenfalls eine Eichung (früher Beglaubigung). Auch hierbei ist Voraussetzung, daß die Bauart des Normals zur Eichung zugelassen ist.

Meßmittelprüfung. Die Eichung ist eine spezielle Form von Meßmittelprüfungen. Durch diese wird ganz allgemein festgestellt, ob ein Meßmittel festgelegte Bedingungen, speziell hinsichtlich der angegebenen Fehlergrenzen, erfüllt. Dazu wird es in angemessenen Fristen mit Referenznormalen verglichen. Unabhängig von den festgelegten *Prüffristen* ist auch dann eine Prüfung durchzuführen, wenn ein begründeter Verdacht besteht, daß von dem betreffenden Meßmittel keine richtigen Werte mehr ausgegeben werden. Für diese Prüfungen sind *Prüfvorschriften* zu erarbeiten. Darin werden die einzelnen Arbeitsschritte wie Vorbereitung, Beschaffenheitsprüfung, Justieren, Meßtechnische Prüfung, Beurkundung sowie die Vorgehensweise und die anzuwendenden Hilfsmittel festgelegt.

7.2.4. Struktur von Meßeinrichtungen

7.2.4.1. Grundstruktur von Meßgeräten. Aufgaben der Meßgeräte. Sieht man von direkten Maßverkörperungen, wie Lineal, Winkelmesser, Hohlmaße usw., ab, so sind durch ein einfaches *anzeigendes Meßgerät* die folgenden drei Aufgaben zu lösen: Durch Ausnutzung einer physikalischen Erscheinung (Meßprinzip) ist ein vom Wert der Meßgröße abhängendes, verarbeitbares Signal zu bilden, in eine geeignete Form zu bringen und auszugeben. Dementsprechend kann man oft drei hintereinandergeschaltete Funktionseinheiten (Meßfühler, Anpasser und Ausgabeeinheit) unterscheiden. Diese Struktur heißt *Meßkette*. Natürlich lassen sich die einzelnen Elemente konstruktiv vielfach nicht trennen; es müssen auch nicht in jedem Fall alle drei vorhanden sein. Eventuell können aber auch noch weitere Funktionseinheiten, insbesondere zusätzliche Anpasser, hinzukommen.

Meßfühler. Der Meßfühler steht unmittelbar oder über eine geeignete Hilfsenergieform, z.B. Strahlung, Luftstrom, mit dem Meßobjekt in Kontakt und erzeugt durch einen Energieumwandlungsprozeß ein *natürliches Abbildungssignal*, das den Wert der Meßgröße eindeutig abbildet. Im einfachsten Fall hängt dieses Abbildungssignal linear vom Wert der Meßgröße ab. Bei einem elektrischen Thermometer, das den thermoelektrischen Effekt ausnutzt, besteht der Meßfühler aus dem Thermopaar. In ihm wird die zugeführte thermische Energie in eine elektrische Spannung umgewandelt, die, in erster Näherung, der Temperaturdifferenz zwischen Meß- und Vergleichsstelle proportional ist. Oft wird der Meßfühler auch als *Aufnehmer* bezeichnet. Der zunehmend verwendete Terminus *Sensor* hat dagegen noch viele andere Bedeutungen.

Anpasser. Der Anpasser bringt das natürliche Abbildungssignal in eine für die Weiterverarbeitung oder Meßwertausgabe geeignete Form. Er kann aus den unterschiedlichsten Funktionselementen bestehen und demzufolge auch mit verschiedenen Bezeichnungen belegt werden. Da die natürlichen Abbildungssignale oft sehr energiearm sind, ist der Anpasser in vielen Fällen ein mechanischer, pneumatischer, optischer, elektrischer usw. Verstärker. Auch Wandler für eine Energieform in eine andere sind nicht selten. Im Beispiel des elektrischen Thermometers ist der Anpasser ein elektrischer Gleichspannungsverstärker, der die sehr kleine Thermospannung in eine gut meßbare Spannung umwandelt. Es gibt viele natürliche Abbildungssignale, die keinen Anpasser benötigen, sondern direkt zur Ausgabe geeignet sind.

Ausgabeeinheit (vgl. 7.2.7.). Das anzeigende Meßgerät gibt den Meßwert in optischer Form aus. Die analoge Ausgabeeinheit muß also eine *Anzeigemarke* und eine *Skale* aufweisen, die relativ zueinander beweglich sind. Von digitalen Ausgabeeinheiten kann man den Zahlenwert in Ziffernform ablesen. Zweckmäßigerweise wird die Skale der Anzeigeeinheit direkt in Einheiten der Meßgröße graduiert. Wenn das nicht möglich ist, wie z.B. bei Vielfachmeßgeräten, müssen die abgelesenen Skalenteile mit Hilfe der Skalenkonstanten in den Meßwert umgerechnet werden. Bei digitaler Anzeige ist dies meist nicht nötig, da beim Umschalten der Meßbereiche die entsprechende Einheit berücksichtigt wird.

7.2.4.2. Meßsysteme. Die einfache Kettenstruktur ist bei Meßeinrichtungen sehr verbreitet, auch wenn sie nicht immer deutlich zu erkennen ist. Ein Nachteil derartiger Anordnungen ist es, daß die zur Messung benötigte Energie

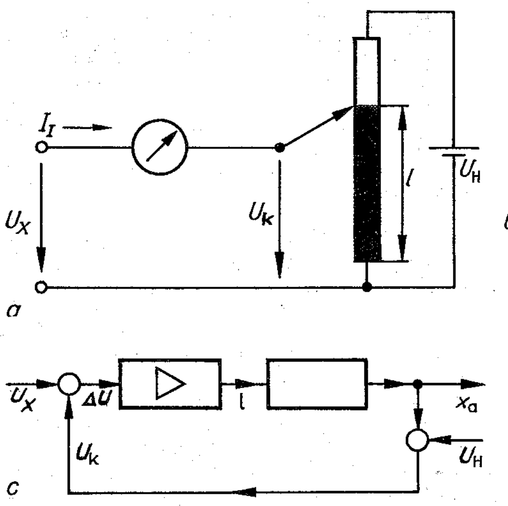
normalerweise dem Meßobjekt entzogen werden muß. So wird beispielsweise bei der Spannungsmessung mit einem Drehspulmeßgerät der Spannungsquelle Strom entnommen und die Spannung dadurch verkleinert. Es kommt also zu einer *Rückwirkung* auf die zu messende Größe. Das ist eine Fehlerursache, denn die eigentlich gesuchte unbelastete Spannung ist größer als der gemessene Wert. Oft ist die Rückwirkung allerdings so klein, daß sie als Fehlerquelle nicht erkennbar wird.

Zur Vermeidung der Rückwirkung u. a. Nachteile der Kettenstruktur wurden die Meßanordnungen weiterentwickelt, und zwar speziell durch die Einführung von Rückführungen. Ein typisches Beispiel dafür sind die *Kompensationsmeßanordnungen*. Zur Erklärung der Wirkungsweise sei ein Motorkompensator für die Messung von elektrischen Spannungen betrachtet (Abb. 7.2.4–1). In der Grundschaltung nach POGGENDORF (Abb. 7.2.4–1a) wird der Meßspannung U_x eine durch ein Potentiometer einstellbare Kompensationsspannung U_K gegengeschaltet. Ist $U_x \neq U_K$, was an dem durch das Indikatormeßgerät fließenden Strom I_I erkennbar ist, so muß das Potentiometer so lange verstellt werden, bis $I_I \to 0$, d. h. $U_x = U_K$. Der Wert von U_K ist an der Potentiometerstellung ablesbar, wenn U_H bekannt und konstant ist. Im Fall $I_I = 0$ wird der Spannungsquelle kein Strom entnommen, U_x wird also leistungslos, d. h. rückwirkungsfrei, gemessen. Das Kompensieren der Meßgröße durch Gegenschalten einer gleich groß zu machenden, entgegengerichteten Kompensationsgröße ist das wesentliche Merkmal der Kompensationsmethode.

Wird nach Abb. 7.2.4–1b die Spannungsdifferenz $\Delta U = U_x - U_K$ einem Verstärker zugeführt, der einen Verstellmotor vorzeichenrichtig ansteuert, über den das Potentiometer nachgestellt wird, so ist der Abgleichvorgang automatisiert. Synchron mit dem Potentiometer wird auch die Meßwertausgabeeinheit verstellt. Damit ist aus der einfachen Kettenstruktur eine Struktur mit *Rückführung* geworden (geschlossene Struktur, Abb. 7.2.4–1c). Durch wiederholte Anwendung der Kompensation (*Mehrfachkompensation*) sowie durch andere Strukturveränderungen können Meßanordnungen immer leistungsfähiger gemacht werden. Insbesondere lassen sich durch geeignete Maßnahmen auch Störeinflüsse verringern. Oft ist es z. B. möglich, durch Zuschalten von Widerständen mit bestimmter Temperaturabhängigkeit den Temperatureinfluß auf elektrische Schaltungen zu verkleinern. Auch Nichtlinearitäten der statischen Kennlinie sind in ähnlicher Weise zu beseitigen oder zumindest zu vermindern. Durch derartige Maßnahmen werden Meßanordnungen immer komplexer. Man spricht dann von *Meßsystemen*, insbesondere, wenn die einzelnen Funktionsgruppen in Form separater Baugruppen oder Einzelgeräte zusammengeschaltet werden. Gerade in jüngster Zeit tritt der Systemcharakter vieler Meßanordnungen zunehmend stärker hervor, indem automatische Steuer- und Rechenfunktionen in Meßgeräte integriert werden (vgl. 7.2.8.), die früher von Hand ausgeführt wurden.

7.2.4.3. Informationstheoretische Betrachtungen. Messen als Informationsaufnahme. Je stärker der Systemcharakter von Meßeinrichtungen in Erscheinung tritt, umso deutlicher sind auch die Analogien zu anderen Einrichtungen der Informationstechnik zu erkennen. Messen bedeutet ja nichts anderes, als Informationen über das Meßobjekt, speziell über den Wert einer Meßgröße zu gewinnen; es ist eine der wichtigsten Formen der Informationsaufnahme. So ist es naheliegend und bei komplizierteren theoretischen Problemen oft sogar notwendig, auch auf die Lösung meßtechnischer Fragen informationstheoretische Methoden anzuwenden. An einigen einfachen Beispielen soll der Zusammenhang zwischen Meßtheorie und Informationstheorie verdeutlicht werden.

Informationsgehalt einer Messung. Die durch eine Messung gewonnene Information ist umso größer, je kleiner die Meßfehler sind, denn bei sehr großen Fehlern erfährt man durch die Messung nur sehr wenig über den richtigen Wert der Meßgröße. Die Anzahl m der unterscheidbaren Meßwertstufen hängt von den Fehlergrenzen des Meßgerätes ab. Bei hundertteiliger Skale und einer Fehlergrenze von $e = \pm 1$ Skt ($e^* = \pm 1\,\%$) sind $m = 51$ Meßwertstufen zu unterscheiden, denn es gilt $m = (2e^*)^{-1} + 1$. Der Summand $+1$ rührt daher, daß an den Meßbereichsgrenzen die unterscheidbaren Stufen nur halbe Breite haben. Sind alle Meßwerte gleichwahrscheinlich, dann ist die Wahrscheinlichkeit p_y für das Auftreten eines Meßwertes $p_y = m^{-1}$. Der *Informationsgehalt* H_0 ist der Wahrscheinlichkeit umgekehrt proportional, und zwar logarithmisch, damit sich Informa-



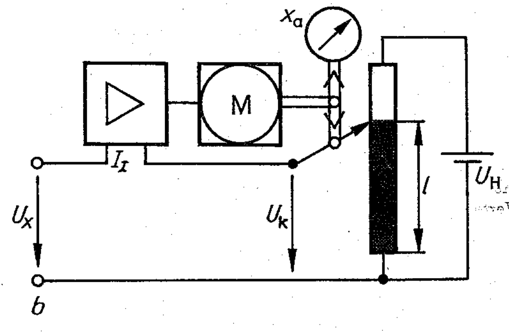


Abb. 7.2.4–1 Spannungskompensation, (a) Grundschaltung nach POGGENDORF, (b) Motorkompensator, (c) Struktur der Kompensationsschaltung

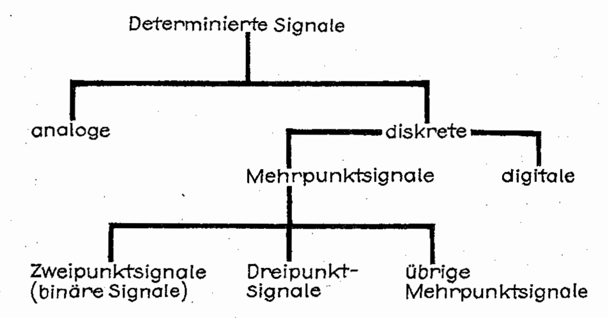


Abb. 7.2.4–2 Einteilung der Signale nach dem Wertevorrat des Informationsparameters

tionsgehalte additiv zusammensetzen lassen und $\sum_{\gamma=1}^{m} p_\gamma = 1$ ist. Daraus folgt $H_0 = \text{lb}\,(p_\gamma^{-1})$, d. h., $H_0 = \text{lb}\, m = \text{lb}\,([2/e^*]^{-1} + 1)$. Somit wird durch ein Meßgerät mit einer Fehlergrenze von 0,5 % bei einer Messung ein Informationsgehalt von $H_0 = \text{lb}\,(0{,}01^{-1} + 1) = \text{lb}\,101$, also 7 bit, vermittelt.

Informationsmenge. Sind die einzelnen Meßwerte – wie es in der Praxis normalerweise der Fall ist – nicht gleichwahrscheinlich, so interessiert der *mittlere Informationsgehalt H*. Er ist die Summe der Informationsgehalte der Einzelwerte unter Berücksichtigung ihrer Wahrscheinlichkeit p_γ, also $H = \sum_{\gamma=1}^{m} p_\gamma H_\gamma = \sum_{\gamma=1}^{m} p_\gamma \text{lb}\, p_\gamma^{-1}$. Handelt es sich um Aufgaben, bei denen zeitlich veränderliche Größen zu messen sind, so müssen viele Meßwerte hintereinander aufgenommen werden, und zwar nach dem Abtasttheorem mindestens doppelt so viele, wie die Meßgröße in der gleichen Zeit Schwingungen ausführt (bei Fourier-Zerlegung). Werden n Meßwerte aufgenommen, so ist das eine *Informationsmenge* von $M = nH$. Die auf die Zeit bezogene Informationsmenge ist der *Informationsfluß* $\dot{M}$, und der maximal mögliche Informationsfluß, der einen Informationskanal passieren kann, ist dessen *Kanalkapazität* C_t; es gilt $C_t \geqq \dot{M}$.

Signale. In der Technik nennt man die Träger von Informationen Signale. Das sind zeitlich veränderliche physikalische Größen, *Signalträger*, von denen ein Parameter, der *Informationsparameter*, die signalisierte Information abbildet. Bei einer Wechselspannung können z. B. Amplitude, Frequenz und Phasenlage als Informationsparameter genutzt werden. Dementsprechend spricht man von amplituden-, frequenz- und phasenmodulierten Signalen. Signale können kontinuierlich und diskontinuierlich sein, d.h., sie sind entweder dauernd oder nur zu bestimmten Zeitpunkten vorhanden. Neben dieser Unterscheidung nach der zeitlichen Verfügbarkeit gibt es noch eine andere Einteilung nach dem Wertevorrat des Informationsparameters (Abb. 7.2.4–2). Bei analogen Signalen sind unendlich viele Werte möglich, bei diskreten Signalen nur eine definierte Anzahl. Unter den *Mehrpunktsignalen* spielen meßtechnisch vor allem Zwei- und Dreipunktsignale eine Rolle. Meßgeräte mit Zweipunktausgangssignalen, d.h. mit Ja/Nein-Aussage, heißen *Schranken* oder *Wächter*, z. B. Füllstandsschranken oder Strömungswächter.

Digitale Signale sind diskrete Signale, bei denen jeder diskrete Wert der signalisierten Größe, d.h. in der Meßtechnik jeder Wert der Meßgröße, mit Hilfe eines Codes in ein Codewort umgesetzt worden ist. Zwischen diesen Codewörtern und den signalisierten Werten muß ein eindeutiger Zusammenhang bestehen. Technisch interessant sind nur *binäre Codewörter*, weil sie sich besonders leicht realisieren lassen. Dabei kann jedes Zeichen des Codewortes nur zwei Werte, nämlich 1 und 0, annehmen, was beispielsweise durch „Strom" und „kein Strom" technisch leicht verwirklicht werden kann. Ein einfacher Code ist z. B. das System der *Dualzahlen*, d. h. die Reihe der Zahlen zur Basis 2. Bei ihnen haben die einzelnen Ziffern von rechts nach links gelesen nicht die Wertigkeiten 10^0, 10^1, 10^2 usw., sondern die Wertigkeiten 2^0, 2^1, 2^2 usw. Danach werden also die Dezimalzahlen von 1 bis 10 als Dualzahlen 1, 10, 11 100, 101, 110, 111, 1 000, 1 001 und 1 010 geschrieben. Wird jede Ziffer einer Dezimalzahl einzeln binär codiert, so ergibt sich ein sog. *bcd-Code* [engl., *b*inary *c*oded *d*ecimal]. Darüber hinaus werden noch verschiedene andere Codes technisch genutzt.

7.2.4.4. Analoge und digitale Meßtechnik. Entsprechend der großen Bedeutung analoger und digitaler Signale lassen sich auch zwei grundsätzlich verschiedene Techniken der Signalübertragung und -verarbeitung unterscheiden, die Analog- und die Digitaltechnik. Das gilt für die Meßtechnik gleichermaßen. In der *analogen Meßtechnik* sind alle Abbildungs- und Zwischensignale analog. Die einzelnen Werte werden also durch Spannungen, Stromstärken, Strecken (Anzeigemarkenausschläge), Drehwinkel usw. dargestellt, bei denen jeder beliebige Zwischenwert innerhalb eines gegebenen Bereiches möglich ist (unendlicher Wertevorrat). Auch die Meßwertausgabe erfolgt durch eine Strecke L oder einen Winkel α. Beim Ablesen (Bilden des Zahlenwertes, vgl. 7.2.3.3.) ist in der Praxis der theoretisch unendlich große Wertevorrat logischerweise nicht auszunutzen, da auf einer Skale kaum mehr als 100 Skt unterzubringen sind. Selbst wenn man durch Interpolation noch jeweils 10 Zwischenwerte bilden könnte, ergäbe das erst einen praktisch nutzbaren Wertebereich der analogen Anzeige von maximal 1 000.

Für manche Zwecke wird auch die Frequenz als Informationsparameter benutzt; man spricht dann von *frequenzanaloger Meßtechnik*. Da die Frequenzmessung vielfach auf einen Zählvorgang zurückgeführt wird (Schwingungen je Zeit), lassen sich unter Umständen hohe Genauigkeiten erreichen. Außerdem ist die Umsetzung der Zählresultate in digitale Signale relativ einfach, so daß die frequenzanaloge Technik eine Zwischenstellung zwischen analoger und digitaler Technik einnimmt.

Bei der *digitalen Meßtechnik* treten in den Meßeinrichtungen – zumindest teilweise – digitale Signale auf. Auch zählende Geräte werden den digitalen zugerechnet. Die Meßwertausgabe erfolgt bei digitalen Meßgeräten im Normalfall ebenfalls digital, d. h. in Form von Ziffern. Damit wird der diskrete Charakter besonders deutlich, denn durch eine dreistellige *Ziffernanzeige* können eben nur 1 000 Werte (0

bis 999) angezeigt werden. Um eine dreieinhalbstellige Anzeige handelt es sich, wenn die erste von vier Stellen nur 1 oder 0 annehmen kann, so daß die Zahlen von 0 bis 1 999 darstellbar sind. Von rein zählenden Geräten abgesehen, ist die erste Wandlung, die Bildung des primären Abbildungssignales, auch in einem digitalen Gerät i. allg. ein analoger Vorgang, durch den oft die erreichbare Genauigkeit bestimmt wird.

7.2.5. Kenngrößen von Meßgeräten

7.2.5.1. Statische Kenngrößen analoger Meßgeräte. Bedeutung statischer Kenngrößen. Zur Beschreibung der meßtechnischen Eigenschaften von Meßmitteln benutzt man bestimmte, in ihrer Bedeutung festgelegte Kenngrößen. Sie beziehen sich entweder auf Eigenschaften, die von zeitlichen Meßgrößenänderungen unabhängig sind (*statische Kenngrößen*) oder auf die Eigenschaft des Meßmittels, zeitlichen Meßgrößenänderungen mehr oder weniger gut folgen zu können (*dynamische Kenngrößen*, vgl. 7.2.5.3.). Die Zahlenwerte der Kenngrößen konkreter Meßmittel heißen Kennwerte, ihre experimentelle Bestimmung ist eine *Kennwertermittlung*.

Meßbereich. Der Wertebereich der Meßgröße, der von einem Meßgerät mit einer Meßunsicherheit gemessen werden kann, die der angegebenen Fehlerklasse (vgl. 7.2.5.4.) entspricht, heißt Meßbereich (MB). Er wird also in Einheiten der Meßgröße angegeben. Ein Außenthermometer kann z. B. einen Meßbereich von −40 °C bis +40 °C haben. Für dieses Thermometer ist −40 °C der *Meßbereichsanfangswert* und +40 °C der *Meßbereichsendwert*. Der *Anzeigebereich* (AB) ist der gesamte Änderungsbereich der Meßgröße, der von einem Gerät angezeigt werden kann, d. h. unabhängig von den dabei möglicherweise auftretenden Fehlern. Oft ist er gleich dem Meßbereich; allgemein gilt AB ≧ MB. Die Differenz zwischen Meßbereichsendwert und -anfangswert ist die *Meßspanne* (MSp). Das genannte Thermometer hat also eine Meßspanne von 80 K, genauso wie ein Thermometer mit einem Meßbereich von 0 bis 80 °C. Die Angabe des Meßbereiches schließt also im Gegensatz zur Meßspanne die Lage auf der Werteskala ein.

Empfindlichkeit. Die Empfindlichkeit S eines anzeigenden Meßgerätes ist die Änderung der Anzeige ΔL, die einer bestimmten Meßgrößenänderung ΔG entspricht, also $S = \Delta L/\Delta G$. Diese Definition läßt sich auch auf die einzelnen Glieder einer Meßkette, die *Übertragungsglieder*, übertragen, die dadurch gekennzeichnet sind, daß zu einer be-

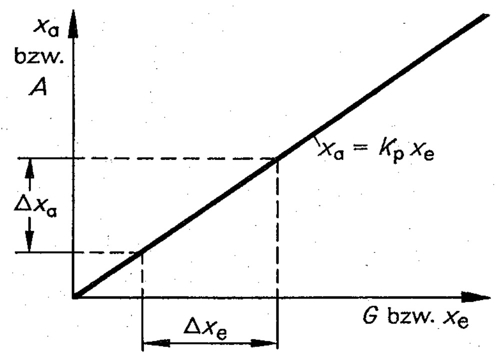


Abb. 7.2.5–1 Beispiel für eine lineare statische Kennlinie

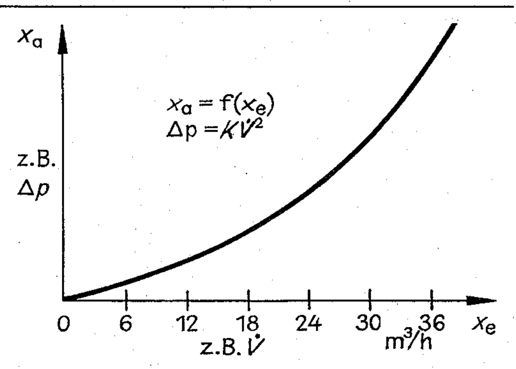


Abb. 7.2.5–2 Statische Kennlinie eines Volumenstrommeßgerätes (Blendenmessung) als Beispiel für nichtlineares Verhalten

stimmten Eingangsgröße x_e auch eine bestimmte Ausgangsgröße x_a gehört. Das i-te Übertragungsglied hat also die Empfindlichkeit $S_i = \Delta x_{ai}/\Delta x_{ei} = \Delta x_{ai}/\Delta x_{a(i-1)}$. Die Maßeinheit der Empfindlichkeit ist $[S] = [x_a]/[x_e]$. Bei einem anzeigenden Voltmeter ist also die Empfindlichkeit z. B. in Zentimeter je Volt anzugeben.

Die Gesamtempfindlichkeit einer Meßkette ist durch die Empfindlichkeiten ihrer einzelnen Übertragungsglieder bestimmt. Die Empfindlichkeiten der drei Übertragungsglieder des elektrischen Thermometers sind also

- Meßfühler (Thermoelement): $S_{TE} = \Delta U_\varphi/\Delta(\Delta\varphi)$ in V/K);
- Anpasser (Gleichspannungsverstärker): $S_V = \Delta U_V/\Delta U_\varphi$ in V/mV;
- Ausgabeeinheit (Voltmeter): $S_A = \Delta L/\Delta U_V$ in cm/V.

Durch Multiplikation der Teilempfindlichkeiten erhält man die *Gesamtempfindlichkeit*, im Beispiel also

$$S_{TE} \cdot S_V \cdot S_A = \frac{\Delta U_\varphi}{\Delta(\Delta\varphi)} \cdot \frac{\Delta U_V}{\Delta U_\varphi} \cdot \frac{\Delta L}{\Delta U_V} = \frac{\Delta L}{\Delta(\Delta\varphi)} = S_{ges} \text{ in cm/K.}$$

Für jede Meßkette gilt $\prod_{i=1}^{r} S_i = S_{ges}$.

Bei hohen Empfindlichkeiten muß man kleine Meßbereiche in Kauf nehmen; benötigt man einen großen Meßbereich, so wird die Empfindlichkeit zwangsläufig kleiner. Deshalb ist, bei Vielfachmeßgeräten, die Angabe des kleinsten Meßbereiches eine Information über die größte Empfindlichkeit des betreffenden Gerätes. Um auch bei hohen Empfindlichkeiten zu größeren Meßbereichen zu kommen, gibt es nur die Möglichkeit, sehr lange Skalen zu benutzen, bei Galvanometern z. B. eine Projektionsskale mit Lichtzeiger.

Ansprechschwelle (Meßschwelle) ist die kleinste Meßgrößenänderung $\Delta x_{e,\,min}$, die eine nutzbare, also ablesbare oder weiterverarbeitbare, Ausgangsgrößenänderung $\Delta x_{a,\,min}$ verursacht. Was noch nutzbar ist, hängt vom Verhältnis Nutzsignal/Rauschen ab und kann beim gleichen Meßgerät für verschiedene Meßgrößen unterschiedlich sein. Die Entscheidung über die Nutzbarkeit kann nur hinsichtlich einer Ausgangsgrößenänderung gefällt werden. Ist also eine Änderung $\Delta x_{a,\,min}$ gerade noch ausnutzbar, so beträgt die Ansprechschwelle $\Delta x_{e,\,min} = S^{-1} \cdot \Delta x_{a,\,min}$, d. h. für empfindliche Geräte kann man auch kleine Meßschwellen erwarten. Bei der Messung von Zeiten und geometrischen Größen bezeichnet man die Ansprechschwelle auch oft als (zeitliches und räumliches) *Auflösungsvermögen*.

Statische Kennlinie nennt man die Darstellung der Ausgangsgröße, z.B. der Anzeige A, als Funktion der Eingangsgröße (der Meßgröße G). Bei linearer Abhängigkeit (Abb. 7.2.5–1) ist also $x_a = K_p x_e$ (z.B. $A = K_p G$). Der *Übertragungsfaktor* K_p charakterisiert ein proportionales Übertragungsverhalten des betreffenden Meßmittels. Im allgemeinen Fall $x_a = f(x_e)$ können unterschiedliche, also auch nichtlineare, Kennlinien auftreten (Abb. 7.2.5–2). Bei anzeigenden Geräten ist ein nichtlineares Verhalten sofort an der Skalenteilung zu erkennen. Normalerweise ist *Proportionalverhalten* für Meßgeräte erstrebenswert, insbesondere wenn die Meßwerte nicht nur auszugeben, sondern auch zu verarbeiten sind. Soll jedoch nur die Einhaltung vorgegebener Sollwerte im letzten Drittel des Meßbereiches kontrolliert werden, so kann z. B. die in Abb. 7.2.5–2 vorliegende *Nichtlinearität* durchaus vorteilhaft sein, weil dann die ersten zwei Drittel des Meßbereiches nur der Orientierung dienen.

Die Steigung der statischen Kennlinie ist die *Empfindlichkeit*. Bei linearem Verhalten gilt das exakt, denn $S = \Delta x_a/\Delta x_e = dx_a/dx_e = K_p$. Überträgt man diese Überlegung auf Übertragungsglieder mit nichtlinearer Kennlinie, so sieht man, daß dort die Empfindlichkeit $S = dx_a/dx_e$ innerhalb des Meßbereiches nicht konstant ist, sondern vom Wert der Meßgröße abhängt: $S = f(x_e)$. Im Fall der Abb. 7.2.5–2 ist demnach die Empfindlichkeit für den interessierenden Meßbereichsabschnitt besonders groß.

Die Gleichung der statischen Kennlinie $x_a = f(x_e)$ ergibt sich aus dem physikalischen Meßprinzip.

Meßbedingungen. Die für ein Meßmittel vom Hersteller zugesicherten Kennwerte gelten natürlich nicht uneingeschränkt, sondern nur bei Einhaltung bestimmter Gebrauchsvorschriften und Umweltbedingungen. Für Geräte mit elektromechanischen Meßwerken wird z.B. angegeben, in welcher Lage, senkrecht, schräg oder waagerecht, sie betrieben werden müssen. Für Geräte mit elektrischer Versorgungsenergie wird eine bestimmte Stromart, Spannung und Frequenz gefordert. Bei mechanischen und optischen Geräten muß die Justage bestimmten Anforderungen genügen.

Die Vorschriften bezüglich der Umgebungsbedingungen beziehen sich generell auf Temperatur, Luftdruck und -feuchte, z. B. +20 °C, 1013 hPa und 65 % relative Luftfeuchte. Darüber hinaus kann noch die Frage elektrischer Felder und mechanischer Schwingungen sowie anderer Erschütterungen eine Rolle spielen. Die geforderten Werte werden i. allg. mit bestimmten Toleranzgrenzen angegeben. So darf bei Betriebsmeßgeräten die Temperatur um ± 5 K um den Nennwert +20 °C schwanken; bei Geräten mit Fehlerklassen < 1,0 verkleinern sich die zulässigen Schwankungen auf ± 2 K, wenn die garantierten Fehlergrenzen eingehalten werden sollen.

Zu den Meßbedingungen zählen schließlich auch die Meßobjekteigenschaften, wie Oberflächenbeschaffenheit, Reinheit, Aggressivität bei Flüssigkeiten und Gasen usw. Auch sie dürfen nur in bestimmten Grenzen schwanken.

7.2.5.2. Statische Kenngrößen digitaler Meßgeräte. Die in 7.2.5.1. genannten Kenngrößen lassen sich sinngemäß auch auf digitale Geräte anwenden. Deshalb genügt es, auf einige Modifikationen hinzuweisen. Die Definition des Meßbereiches gilt bei den eigentlichen digitalen Meßgeräten, z. B. Digitalvoltmetern, unverändert, Zähler dagegen werden durch den *Belastungsbereich* charakterisiert. Das ist der Wertebereich der Meßgröße, innerhalb dessen der durch Summation gebildete Meßwert die Fehlergrenzen einhält. Dieser abweichende Terminus ist deswegen notwendig, weil Meßwert, z.B. Volumen V, und Eingangsgröße, z.B. Volumenstrom $\dot{V}$, von unterschiedlicher Größenart sind. Der Belastungsbereich hängt von der Zählgeschwindigkeit, die *Kapazität* des Zählers (der größtmögliche Meßwert) von der Stellenzahl der Anzeige ab. Die Grenzen des Belastungsbereiches sind die obere und untere *Belastungsgrenze*. Bei mechanischen Zählern ist außerdem noch der *Anlaufwert* zu berücksichtigen. Das ist der Wert der Meßgröße, bei dem der Zähler überhaupt erst zu laufen beginnt.

Die *Empfindlichkeit* digitaler Meßgeräte ist der kleinste Ziffernschritt ΔZ (ein Schritt der letzten Stelle) bezogen auf die dafür erforderliche Meßgrößenänderung ΔG, also $S_{dig} = \Delta Z/\Delta G$. Daraus folgt als Einheit für S_{dig}: $[S_{dig}] = [G]^{-1}$. Da ΔZ durch Stellenzahl und Meßbereich eindeutig festgelegt ist, läßt sich S_{dig} auch durch ΔG ausdrücken. Die Meßgrößenänderung ΔG, die ΔZ entspricht, also $\Delta G = S_{dig}{}^{-1} \times \Delta Z$, ist zugleich das *Auflösungsvermögen* eines digitalen Gerätes. Hinsichtlich der Meßbedingungen gibt es im Vergleich zu den analogen Geräten in der digitalen Meßtechnik keine Besonderheiten. Im allgemeinen werden die gleichen Umweltforderungen gestellt wie bei allen anderen elektronischen Meßgeräten.

7.2.5.3. Dynamische Kenngrößen. Zeitverhalten von Meßgeräten. Die dynamischen Kenngrößen eines Meßgerätes geben darüber Auskunft, wie es auf zeitliche Meßgrößenänderungen reagiert. Mathematisch läßt sich das Zeitverhalten eines Meßgerätes durch seine Differentialgleichung beschreiben. Für den Fall der Federwaage lautet die Differentialgleichung der Meßanordnung also $m\ddot{s} + K_R\dot{s} + D_F s = mg$ (K_R = Reibungskoeffizient). Da als höchste zeitliche Ableitung $\ddot{s}$ auftritt, handelt es sich um eine Verzögerung zweiter Ordnung. Die Aufstellung, Lösung und auch Deutung der Differentialgleichung wird bei Verzögerungen höherer Ordnung so kompliziert, daß man für die Beschreibung des dynamischen Verhaltens realer Meßgeräte Kenngrößen benutzt, die sich experimentell ermitteln lassen. Je nachdem, ob man zur Bestimmung dynamischer Kennwerte eine einmalige zeitliche Änderung der Eingangsgröße oder periodische Testsignale unterschiedlicher Frequenz benutzt, unterscheidet man dynamische Kenngrößen im Zeitbereich und im Frequenzbereich.

Kenngrößen im Zeitbereich. Ändert sich die Meßgröße am Eingang eines Meßgerätes beispielsweise sprungförmig von einem niederen auf einen höheren Wert (Abb. 7.2.5–3), so folgt die Anzeige normalerweise nicht verzögerungsfrei. Wird ein 20 °C warmes Thermometer plötzlich in siedendes Wasser gesteckt und verfolgt, wie sich die angezeigte Temperatur mit der Zeit ändert, so kann sich der gezeigte Verlauf ergeben. Diese Kurve heißt entsprechend der sprungförmigen Eingangsgrößenänderung *Sprungantwort* bzw., bei Normierung auf die Sprunghöhe 1, Einheitssprungantwort oder *Übergangsfunktion* $h(t)$. Die abgebildete Übergangsfunktion gehört zu einem Meßgerät mit Verzögerung erster Ordnung. Die Zeit, die vergeht, bis diese Kurve 63,2 % des Endwertes im eingeschwungenen Zustand erreicht hat, ist die Zeitkonstante T_1 des Gerätes. T_1 ist die Konstante bei der ersten Ableitung in der Differentialgleichung, die das Zeitverhalten beschreibt. Die Zeitspanne von t_0 bis zu dem Zeitpunkt, wo die Übergangsfunktion nur noch ε % vom endgültigen Wert abweicht, ist die *Beruhigungszeit* T_ε, früher auch Einschwing- oder Einstellzeit genannt. Meist wird ε mit 5 % angesetzt.

Bei Verzögerung höherer Ordnung treten mehrere Zeitkonstanten auf, deren Bestimmung komplizierter ist, besonders wenn sie alle unterschiedliche Werte haben. Ein Beispiel

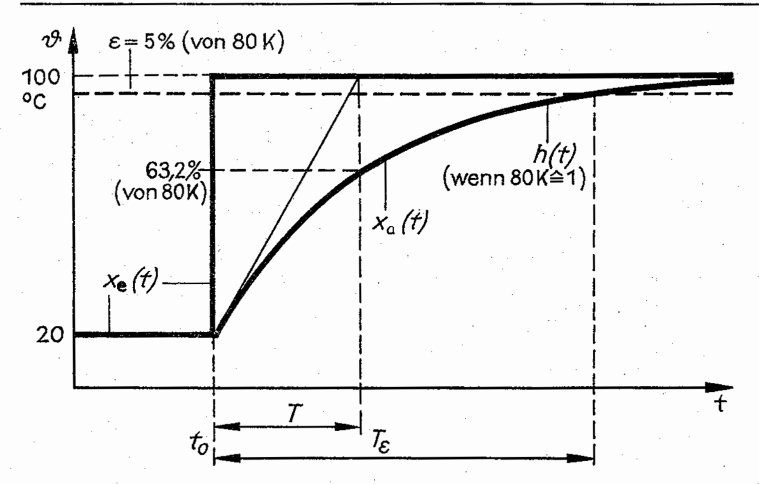


Abb. 7.2.5–3 Zeitverlauf von Eingangs- und Ausgangsgröße eines Thermometers bei sprungförmiger Temperaturänderung

für die Übergangsfunktion eines schwingungsfähigen Meßgerätes zeigt Abb. 7.2.5–4. Daß es sich nicht um eine Verzögerung erster Ordnung handelt, erkennt man am Vorhandensein eines Wendepunktes bei $h(t)$. Mit Hilfe einer Tangente an den Wendepunkt lassen sich u. a. Ordnung der Verzögerung und Zeitkonstanten ermitteln. Die Beruhigungszeit T_ε läuft bis zu dem Moment, wo die Toleranzgrenzen $\pm\varepsilon$ um den endgültigen Wert nicht mehr überschritten werden (vgl. Abb. 7.2.5–4). Aus der Periodendauer des Einschwingvorganges T_p kann die Eigenfrequenz ω_0, aus den Überschwingweiten h_i die Dämpfung des Meßgerätes errechnet werden.

Kenngrößen im Frequenzbereich. Liegt am Eingang eines Meßgerätes eine Meßgröße an, deren Wert sich periodisch, z. B. sinusförmig, ändert, so wird sich im eingeschwungenen Zustand auch die Ausgangsgröße (Anzeige) sinusförmig ändern. Wichtige Kenngrößen sind hier das Amplitudenverhältnis $|G| = |x_{a0}| / |x_{e0}|$ der beiden Schwingungen, die x_e und x_a ausführen, und die Phasenverschiebung φ (Abb. 7.2.5–5). Der Logarithmus von $|G|$ über $\lg(\omega/\omega_0)$ aufgetragen ist die *Amplitudenkennlinie*, φ über $\lg(\omega/\omega_0)$ ergibt die *Phasenkennlinie* des betreffenden Meßgerätes. Da bei diesen Kurven ω anstelle von t als Abszisse auftaucht, handelt es sich um Betrachtungen im *Frequenzbereich*. Durch weitere Überlegungen und Umformungen kommt man weiterhin zum *Frequenzgang* und zur *Übergangsfunktion*, die das dynamische Verhalten ebenfalls umfassend beschreiben.

Während bei einem Gerät mit idealem Zeitverhalten das Amplitudenverhältnis $|G| = 1$ ist, d. h., eine Schwingung von diesem Meßgerät ungedämpft abgebildet wird, läßt sich für reale Meßgeräte aus der Frequenzabhängigkeit von $|G|$, also aus $|G| = f(f)$, die Frequenz ablesen, bis zu der das Meßgerät noch brauchbar ist. Nimmt man eine bestimmte Dämpfung als noch zulässig an, z. B. bis zu einem Wert $|G| = 1/\sqrt{2}$, so ist die Frequenz, bei der die Dämpfung diesen zulässigen Wert überschreitet, die *Grenzfrequenz* f_g. Manche Meßgeräte, wie piezoelektrische Druckmeßfühler oder Geräte mit Wechselspannungsverstärkern, funktionieren erst bei einer Mindestfrequenz der Meßgröße; sie haben *Bandpaßverhalten* und demzufolge eine untere und obere Grenzfrequenz f_{gu} und f_{go}. Die vorgenannten Geräte hatten *Tiefpaßverhalten*. Bei diesen Geräten besteht zwischen f_g und der Beruhigungszeit T_ε der Zusammenhang $2\,T_\varepsilon = 1/f_g$. Diese Beziehung zeigt, daß die Kenngrößen im Zeit- und Frequenzbereich nur unterschiedliche Beschreibungen des gleichen Sachverhaltes darstellen.

7.2.5.4. Fehlerkenngrößen von Meßgeräten. Technische Unzulänglichkeiten und unvermeidliche Störeinflüsse führen dazu, daß ein Meßgerät den Wert einer Meßgröße nicht

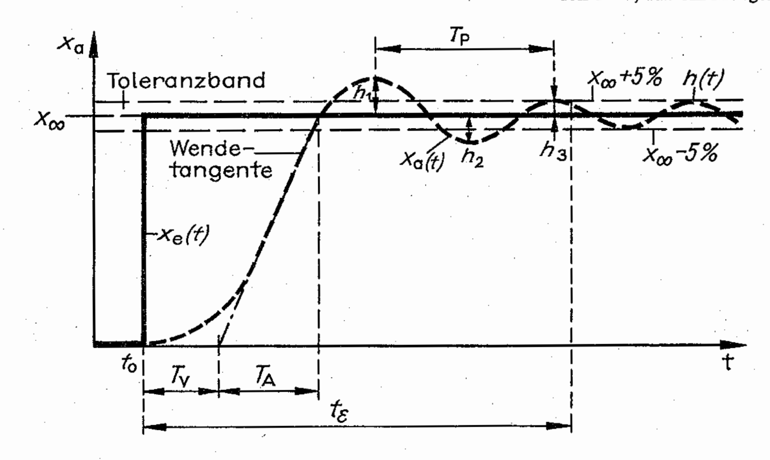


Abb. 7.2.5–4 Übergangsfunktion eines schwingungsfähigen Meßgerätes (aus T_A und T_V können die Zeitkonstanten ermittelt werden)

völlig richtig anzeigt. Ist diese Abweichung des Meßwertes vom richtigen Wert der Meßgröße nach Betrag und Vorzeichen konstant, so handelt es sich um einen systematischen Fehler (vgl. 7.2.6.2.). Solche Fehler werden bei der Prüfung des Meßgerätes (vgl. 7.2.3.4.) erkannt und müssen entweder behoben oder durch entsprechende Korrekturen berücksichtigt werden. Darüber hinaus gibt es aber noch zufällige Schwankungen der Anzeige, die auf verschiedene Ursachen zurückgehen. Diese und die nicht erkannten bzw. wegen ihrer Kleinheit eventuell vernachlässigten systematischen Restfehler machen die Anzeige eines Meßgerätes mehr oder minder unsicher. Diese Tatsache wird durch die Angabe von *Fehlergrenzen* berücksichtigt. Dabei kann es sich um Grenzen handeln, die vom Gerätehersteller garantiert werden (Garantiefehlergrenzen), um solche, die für bestimmte Anwendungsfälle behördlich gefordert werden (z. B. Eichfehlergrenzen), oder um Fehlergrenzen, die ein Gerät wirklich einhält. Diese sind für den Meßgeräteanwender besonders wichtig, da sie unabhängig von den einzelnen Fehlerursachen Aussagen über den möglichen Gesamtfehler machen. Sie heißen *Grundfehlergrenzen* (abkürzend auch als *Grundfehler* bezeichnet). In ihnen sind u. a. die durch Linearitätsabweichungen, unzureichendes Auflösungsvermögen, auftretende Drift und die durch das Auftreten von Hysterese verursachten Fehleranteile berücksichtigt.

Die Grundfehlergrenzen werden nur eine bestimmte Zeit lang, (den Gültigkeitszeitraum), eingehalten. Außerdem gelten sie nur für die vorgeschriebenen Meßbedingungen. Werden die Toleranzgrenzen für die Nennbedingungen von Temperatur, Luftdruck usw. überschritten, so können weitere *Zusatzfehler*, z. B. ein Temperaturzusatzfehler von 0,01 %/K, auftreten.

Aus der in Prozent berechneten, auf die Meßspanne bezogenen Grundfehlergrenze wird die *Fehlerklasse*, auch Genauigkeitsklasse, Klassengenauigkeit oder abkürzend Klasse, des Gerätes abgeleitet. (Bei Längenmeßgeräten gilt eine abweichende Definition für die Fehlerklasse.) Für elektrische Meßgeräte sind z.B. die folgenden Vorzugswerte für die anzuwendenden Fehlerklassen festgelegt:

0,05; 0,1; 0,2; 0,5; 1; 1,5; 2,5 und 5.

Die Aussage, ein Gerät habe die Fehlerklasse 1,5, bedeutet, daß seine Grundfehlergrenze $> \pm 1\,\%$ ist, aber $\pm 1{,}5\,\%$ nicht überschreitet. Die Fehlerklasse gilt, wenn nicht ausdrücklich etwas Abweichendes angegeben ist, im ganzen Meßbereich.

7.2.6. Fehler bei Messungen

7.2.6.1. Meßfehler und ihre Einheiten. Bei allen Messungen ist davon auszugehen, daß der gemessene Wert x_m vom gesuchten richtigen Wert x_r der Meßgröße mehr oder weniger abweicht. Diese *Abweichung* bezeichnet man üblicherweise als *Meßfehler e* oder kurz als *Fehler*. Es gilt also $e = x_m - x_r$. Ein zu großer Meßwert ist folglich mit einem positiven und ein zu kleiner Meßwert mit einem negativen Meßfehler behaftet.

Der Meßfehler e hat die gleiche Einheit wie die Meßgröße; bei der Messung eines elektrischen Widerstandes ist also $[e] = \Omega$. Deswegen heißt *e absoluter Fehler*.

Absolute Fehlerangaben sind in vielen Fällen unzweckmäßig, z. B. weil sie schwer vergleichbar sind. Deshalb bezieht man den Fehler einer Messung oft auf den Wert der Meßgröße und kommt so zum *relativen Fehler* e^*. Da der richtige

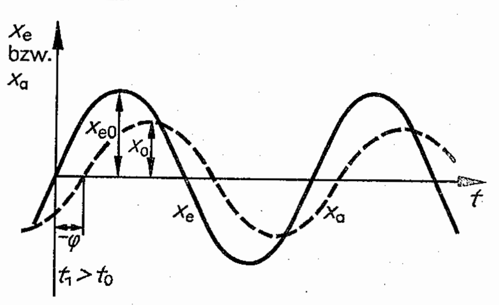


Abb. 7.2.5–5 Sinusförmige Testfunktion x_e und die ihr entsprechende Antwortfunktion x_a im eingeschwungenen Zustand eines Meßgerätes mit Proportionalverhalten und Verzögerung

Wert bei einer Messung nicht bekannt ist, darf man auch den Meßwert als Bezugswert verwenden, also $e^* = e/x_m$. Soll der relative Fehler nicht als Quotient, sondern in Prozent angegeben werden, so gilt $e^*_{\%} = e/x_m \cdot 100$.

Bei Meßmitteln benötigt man einen anderen Bezugswert, da mit ihnen die verschiedensten Meßwerte gemessen werden. Besonders geeignet ist statt x_m die Meßspanne *MSp*. Oft wird auch der Meßbereichsendwert empfohlen, was bei Meßgeräten mit natürlichem Nullpunkt, d.h., Meßwert 0 ist gleich Meßbereichsanfangswert, auf dasselbe hinausläuft. Der so definierte Fehler, der auch zur Ableitung der Grundfehlergrenze (vgl. 7.2.5.3.) benutzt wird, heißt *reduzierter Fehler* e^0 (lies: *e*-Kreis), also $e^0 = e/MSp$, und bei Angabe in Prozenten $e^0_{\%} = 100\, e/MSp$. Die Kenntnis der Fehlerklasse und damit des reduzierten Fehlers gibt die Möglichkeit, den relativen Fehler einer mit dem betreffenden Gerät durchgeführten Messung in Prozent zu bestimmen, nämlich $e^* = e^0\, MSp/x_m$. Mißt man z.B. mit einem Stromstärkemeßgerät der Fehlerklasse 0,5, das einen Meßbereich von 0 bis 6 A hat, den Wert $x_m = 4$ A, so ist der relative Fehler der Messung $e^* = 0{,}5\,\% \cdot 6\,\mathrm{A}/4\,\mathrm{A} = 0{,}75\,\%$. Er beträgt bei 4 A also bereits das Anderthalbfache des aus der Fehlerklasse abgeleiteten reduzierten Fehlers. Deshalb ist zu empfehlen, ein Meßgerät möglichst so auszuwählen, daß die Meßwerte im oberen Drittel des Meßbereiches liegen.

7.2.6.2. Arten von Meßfehlern. Einteilungsmöglichkeiten. Der Meßfehler e setzt sich aus mehreren Bestandteilen zusammen, die unterschiedliche Ursachen haben. So lassen sich beispielsweise Fehlerbeiträge unterscheiden, die vom Messenden nicht beeinflußbar sind, die *objektiven Fehler*, und Beiträge, die der Messende durch sorgfältiges Vorgehen vermeiden kann, die *subjektiven Fehler*. Im Hinblick auf die notwendigen Fehlerbetrachtungen bei der Auswertung von Messungen ist jedoch eine Unterscheidung in zeitabhängige und zeitlich konstante Fehler (Abb. 7.2.6–1) besonders wichtig. Wie die Abb. zeigt, sind bei beiden Gruppen von Fehlern noch weitere Differenzierungen notwendig, z. B. grobe, systematische und zufällige Fehler. Die Einteilung in verschiedene *Fehleranteile*, die wegen ihrer Bedeutung für die Fehlerrechnung hier im Vordergrund steht, läßt sich anhand von Abb. 7.2.6–2 verdeutlichen.

Durch den *systematischen Anteil* e_s des Fehlers e werden im Beispiel alle Meßwerte um den gleichen Betrag vergrößert, da die Ursachen, die zu systematischen Fehleranteilen führen, auch bei wiederholten Messungen stets in gleicher Größe und in gleicher Richtung (gleiches Vorzeichen) wirksam werden. Ungeachtet dieses systematischen Fehleranteil-

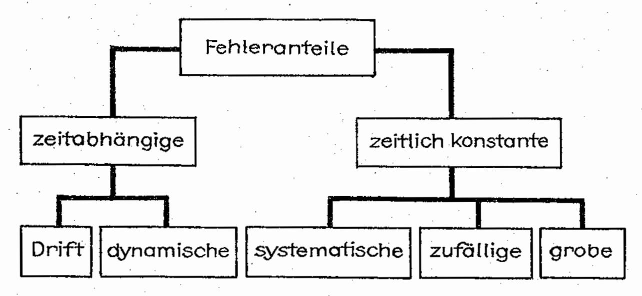


Abb. 7.2.6–1 Übersicht über die möglichen Anteile in einem Meßfehler

les streut jede Einzelmessung außerdem um den Mittelwert $\bar{x}$ einer Meßreihe, und zwar infolge unkontrollierbarer zufälliger Einwirkungen. Damit kommt zu dem systematischen Anteil noch ein nach Betrag und Vorzeichen stochastisch schwankender *zufälliger Fehleranteil* e_f [lat. fortuitus, zufällig] hinzu. So setzt sich also jeder einzelne Meßfehler e_i jeweils aus mehreren Anteilen zusammen, die einzeln untersucht werden müssen, da sie wegen ihres unterschiedlichen Charakters nicht in einem Zahlenwert zusammengefaßt werden können.

Zeitabhängige Fehleranteile sind die dynamischen und die Driftfehleranteile. Unter *dynamischen Fehlern* versteht man die Anteile des Gesamtfehlers, die aus dem nichtidealen dynamischen Verhalten realer Meßgeräte herrühren (vgl. 7.2.5.5.). Es ist leicht einzusehen, daß von einem Gerät mit der in Abb. 7.2.5–3 gezeigten Übergangsfunktion der Meßwert im Zeitpunkt T_1 um 36,8 % zu klein angezeigt wird. Bei einer statischen Messung wird man selbstverständlich warten, bis die Anzeigemarke zur Ruhe gekommen ist, also mindestens bis zum Zeitpunkt T_ε. Dann ist der dynamische Fehleranteil normalerweise zu vernachlässigen. Mißt man jedoch schnell veränderliche Größen, so ist oft schwer erkennbar, wie groß die dynamischen Fehleranteile infolge des trägheitsbedingten „Nachhinkens" des Meßgerätes sind.

Bei periodisch zeitabhängigen Meßgrößen spielen dynamische Fehleranteile gleichfalls eine Rolle. Schon bei Meßgrößenfrequenzen in der Nähe der Grenzfrequenz liegen die möglichen dynamischen Fehleranteile in der Größenordnung von 30 %. Werden aperiodisch zeitabhängige Größen gemessen, so kann man die Grenzfrequenz ebenfalls als Kriterium für das mögliche Auftreten dynamischer Fehler benutzen, da solche Zeitverläufe mit Hilfe der Fourier-Zerlegung auf periodische Vorgänge zurückgeführt werden können. Dabei ist die höchste auftretende Frequenz maßgeblich dafür, ob der Zeitverlauf einer Größe durch ein Meßgerät unverfälscht oder fehlerhaft abgebildet wird. Bei

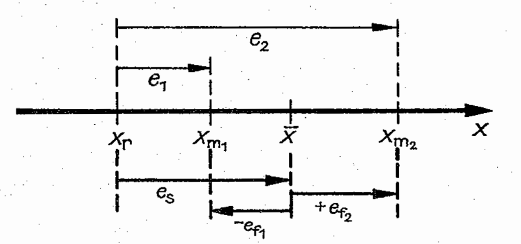


Abb. 7.2.6–2 Zusammensetzung der Meßfehler e_1 und e_2 aus jeweils einem systematischen Anteil e_s und zufälligen Anteilen e_{f_1} und e_{f_2}

allen Messungen zeitabhängiger Größen ist also die Möglichkeit des Auftretens dynamischer Fehleranteile in Erwägung zu ziehen.

Driftfehler sind Fehleranteile, die nicht auf die Zeitabhängigkeit der gemessenen Größe, sondern auf langzeitliche Änderungen von Meßgeräteparametern oder Meßbedingungen zurückzuführen sind. Da sie sich erst nach längerer Zeit (Tage oder Monate) auswirken, sind sie normalerweise sehr schwer zu erkennen. Sie werden beispielsweise durch Alterungserscheinungen von Bauelementen in den Meßgeräten verursacht. Bei Einzelmessungen gibt es keine Driftfehler, sie spielen nur bei der Kontrolle von Meßgrößen über längere Zeiten eine Rolle. Gerade das „Wegdriften" von Meßgeräteparametern, das die Ursache für Driftfehler ist, zwingt zur regelmäßigen Prüfung von Meßmitteln.

Grobe Fehler. Die in Abb. 7.2.6–1 angeführten groben Fehler gehen auf echte Mängel bzw. Defekte der benutzten Meßmittel oder auf falsche Bedienung bzw. Ablesung durch den Messenden zurück. Es handelt sich bei ihnen um Abweichungen, welche die normale Streuung der Meßwerte deutlich überschreiten. Gegen grobe Fehler kann man sich nur durch sorgfältiges Arbeiten und regelmäßige Prüfung der Geräte schützen. Sobald begründete Annahmen über einen Meßmitteldefekt vorliegen, sollte deshalb das betreffende Gerät unabhängig von vorgegebenen Prüffristen durch Vergleich mit einem anderen Meßgerät oder durch Messung von Meßgrößen, deren Werte gut bekannt sind, zumindest grob kontrolliert werden. Wichtig ist es auch, sich vor jeder Messung durch theoretische Überlegungen eine Vorstellung vom ungefähren Wert der zu messenden Größe zu verschaffen, um den Meßwert einer *Plausibilitätsbetrachtung* unterziehen zu können. Damit wird das Erkennen grober Fehler wesentlich erleichtert. Bei Meßreihen werden grobe Fehler als *Ausreißer* bezeichnet. Für das Erkennen von Ausreißern gibt es verschiedene Ausreißerkriterien. Ist beispielsweise $\bar{x}$ der Mittelwert und s die Standardabweichung einer normalverteilten Meßreihe (vgl. 7.2.6.4.) und besteht der Verdacht, daß x_{gr} angesichts einer großen Abweichung vom Mittelwert ein Ausreißer ist, so kann x_{gr} mit einer statistischen Sicherheit P als Ausreißer angesehen werden, wenn $|x_{gr} - \bar{x}| > s\,k_A$ gilt. Der Ausreißerfaktor k_A ist der Tab. 7.2.6–3 zu entnehmen. In den folgenden Fehlerbetrachtungen wird davon ausgegangen, daß keine groben Fehler bzw. Ausreißer aufgetreten sind.

Systematische Fehleranteile sind aufgrund der ständig in gleicher Weise wirkenden Ursachen nach Betrag und Vorzeichen annähernd konstant. Wie Abb. 7.2.6–2 zeigt, sind sie auch durch eine Wiederholung der Messung nicht zu erkennen, und bei Meßreihen sind sie im Mittelwert in vollem Umfang mit enthalten. Etwas simplifizierend kann man additive und multiplikative systematische Fehleranteile unterscheiden. Die *additiven systematischen Fehler* sind

Tab. 7.2.6–3 k_A-Faktoren des Ausreißerkriteriums

n	statistische Sicherheit 95 %	99 %	99,73 %
9	4,42	7,10	11,49
12	4,16	6,38	8,80
15	4,03	5,88	7,66
20	3,90	5,41	6,73
25	3,84	5,14	6,25
30	3,80	5,00	5,95
40	3,75	4,82	5,56
50	3,73	4,70	5,34

über den ganzen Meßbereich konstant. Deshalb sind sie auch beim Meßwert „Null" als Nullpunktverschiebung zu beobachten. Sie werden durch eine *Nullpunktkorrektur* behoben oder bei Geräten, bei denen dies technisch nicht möglich ist, durch eine Korrektion am Meßwert berücksichtigt. Zeigt z. B. ein Druckmeßgerät beim Druck $p = 0$ einen Wert von 3 Pa an und läßt sich eine Nullpunktkorrektur nicht durchführen, so sind von jedem gemessenen Wert 3 Pa abzuziehen, um diesen Fehler zu beseitigen.

Problematischer sind die *multiplikativen systematischen Fehleranteile*. Sie sind an jeder Stelle des Meßbereiches verschieden groß. Der in Abb. 7.2.6–4 gezeigte Verlauf ist noch der einfachste Fall; es gibt auch meßwertabhängige systematische Fehler, die sich innerhalb des Meßbereiches unterschiedlich ändern und sogar das Vorzeichen wechseln. Da systematische Fehler grundsätzlich durch eine *Korrektion* aus dem Meßergebnis zu tilgen sind, muß die Fehlerkennlinie oder -kurve bekannt sein, um für jeden gemessenen Wert die Größe der Korrektion bestimmen zu können. Diese Kurve sollte als Ergebnis der Prüfung für jedes Präzisionsmeßgerät und Normal vorliegen. Oft kommt es in der Praxis vor, daß die Fehlerkurve eines Meßgerätes nicht bekannt ist. Will man sich trotzdem eine Übersicht über die zu erwartenden systematischen Fehleranteile verschaffen, so gibt es prinzipiell die in Abb. 7.2.6–4 gezeigten Möglichkeiten.

Die *Abschätzung systematischer Fehleranteile* setzt eine gute meßtechnische Erfahrung und eine genaue Kenntnis des Meßgerätes und der Meßbedingungen einschließlich ihrer Auswirkungen auf die Messung voraus. Sie ist auf jeden Fall erheblich unsicherer als das experimentelle Verfahren. Es wird nie möglich sein, die systematischen Fehleranteile durch die Korrektion vollständig aus dem Meßergebnis zu tilgen, d. h., es bleibt ein nicht erfaßter systematischer Fehleranteil, der *systematische Restfehler* e_{sR}, übrig. Er hat einen mehr zufälligen Charakter und wird deshalb in der Meßunsicherheit berücksichtigt.

Zufällige Fehleranteile e_f lassen sich nicht exakt erfassen, d. h., sie schwanken nach Größe und Vorzeichen. Deswegen gibt es auch keine Möglichkeit, sie durch eine Korrektion zu berücksichtigen; denn sie sind für den einzelnen Meßwert grundsätzlich unbekannt und machen ihn unsicher. Deshalb werden sie mit statistischen Verfahren abgeschätzt und beim Meßergebnis mit angegeben (vgl. 7.2.6.3.). Der Einfluß zufälliger Fehler wird geringer, wenn eine Messung mehrmals wiederholt und ein Mittelwert $\bar{x}$ als Meßergebnis gebildet wird. Auf die zufälligen Fehler wird in 7.2.6.4. näher eingegangen.

7.2.6.3. Meßunsicherheit in Einzelmeßwerten. Oft wird eine Größe nur einmal gemessen, entweder aus ökonomischen Gründen oder weil sie sich ändert (z. B. beim punktweisen Aufnehmen des Zeitverlaufes der Größe) bzw. die Meßbedingungen nicht konstant gehalten werden können (keine *Wiederholbedingungen*). Es ist auch möglich, daß bei wiederholten Messungen jedesmal der gleiche Wert abgelesen wird, weil die zufälligen Schwankungen kleiner sind als die Ablesegenauigkeit des Meßgerätes, was bei elektrischen Messungen oft der Fall ist. In diesen Fällen liegt also nach der Messung nur ein Wert vor, der jedoch wie jeder andere Meßwert auch mit Fehlern behaftet ist. Für die systematischen Anteile dieser Fehler gilt das unter 7.2.5.2. Gesagte gleichermaßen. Auch hier sind die erhaltenen Meßwerte durch Anbringen der Korrektion ($-e_s$) hinsichtlich der systematischen Fehler zu korrigieren.

Zur Angabe des vollständigen Meßergebnisses gehört in jedem Fall eine Aussage über die Unsicherheit, mit der dieses Meßergebnis behaftet ist oder richtiger, behaftet sein könnte. Als *vollständiges Meßergebnis* ist also stets anzugeben:

– das (hinsichtlich systematischer Fehler) korrigierte Meßergebnis,
– die Meßunsicherheit,
– die statistische Sicherheit für die angegebene Meßunsicherheit.

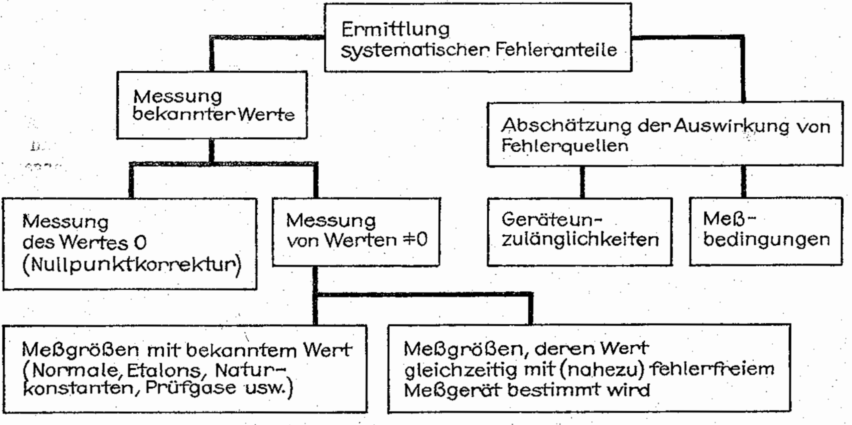


Abb. 7.2.6–4 Möglichkeiten zur Ermittlung systematischer Fehleranteile von Meßgeräten

Die *Meßunsicherheit* u kann wie der Fehler e als absolute oder relative Meßunsicherheit u^* (auch in Prozent) angegeben werden. Das vollständige Meßergebnis einer Kraftmessung mit dem korrigierten Meßwert $F = 75$ N kann also beispielsweise lauten:

$F = (75 \pm 2)$ N; $(P = 95\,\%)$,

$F = 75$ N; $u_{95} = 2$ N,

$F = 75\,(1 \pm 0{,}0267)$ N; $(P = 95\,\%)$,

$F = 75$ N; $u_{95}^* = 2{,}67\,\%$ (bzw. $= 2{,}67 \cdot 10^{-2}$).

Wird aus irgendwelchen Gründen, z. B. in Tab. bei Platzmangel, die Meßunsicherheit nicht mit angegeben, so ist die Anzahl der signifikanten Ziffern des korrigierten Meßergebnisses so zu wählen, daß die Meßunsicherheit klein ist gegenüber dem Stellenwert der letzten Ziffer. Zur Berechnung der Meßunsicherheit u geht man wenn möglich von der Standardabweichung s (vgl. 7.2.6.4.) aus. Bei oft wiederkehrenden Einzelmessungen kann man beispielsweise die Standardabweichung zuvor einmal ermitteln, indem man eine ähnliche Messung unter Wiederholbedingungen mehrmals durchführt und u aus dem so ermittelten s berechnet. Ist dies nicht möglich, z. B. weil alle Ablesungen unter Wiederholbedingungen auf den gleichen Meßwert führen, so nimmt man an, daß die Unsicherheit der Messung nur durch die Fehlergrenzen des Meßgerätes (vgl. 7.2.5.4.) bedingt sind. Man errechnet also nach der Beziehung $e_i = e^0\,MSp$ den absoluten bzw. nach $e_i^* = e^0\,MSp/x_m$ dem relativen Fehler der Messung und bestimmt daraus durch Berücksichtigung des systematischen Restfehlers e_{sR} die Meßunsicherheit.

Die Zusammenfassung des Fehlers der Messung (bzw. der Vertrauensgrenzen, vgl. 7.2.6.4.) mit dem systematischen Restfehler kann durch algebraische bzw. pythagoreische Addition erfolgen, also

$u = e_i + e_{sR}$ bzw. $u = +\sqrt{e_i^2 + e_{sR}^2}$.

Im allgemeinen ist die pythagoreische Addition zu bevorzugen. Diese Frage ist jedoch nahezu gegenstandslos, wenn man bedenkt, wie schwierig und damit auch recht willkürlich die Bestimmung von e_{sR} ist und welche Unsicherheit diesem Wert folglich anhaftet. Wer aufgrund fehlender Erfahrung dazu neigt, e_{sR} zu klein zu schätzen, addiert zweckmäßigerweise algebraisch. Werden für e_i und e_{sR} die relativen Werte eingesetzt, so ergibt sich auch u^* anstelle von u.

7.2.6.4. Meßunsicherheit bei Meßreihen. Wird eine Meßgröße unter Wiederholbedingungen n mal gemessen, so ergeben sich n Meßwerte x_i, die infolge der zufälligen Fehleranteile mehr oder weniger streuen. Handelt es sich um völlig zufällige Fehler, wie sie aufgrund stochastisch wirkender Ursachen zustande kommen, so sind sie voneinander unabhängig (nicht korreliert) und normalverteilt. Die Wahrscheinlichkeit für das Auftreten eines bestimmten zufälligen Fehlers läßt sich demzufolge durch die Gaußsche Fehlerkurve (Glockenkurve) beschreiben:

$$\varphi(x;\mu,\sigma) = \left(\sigma\sqrt{2\pi}\right)^{-1} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right].$$

Abb. 7.2.6–5 zeigt diese Kurve in standardisierter Form, d. h., der Erwartungswert μ ist $= 0$ gesetzt, und die Standardabweichung σ hat den Wert 1 erhalten, also:

$$\varphi(x;0,1) = (2\pi)^{-\frac{1}{2}} \exp(-x^2/2).$$

Aus dieser Kurve läßt sich beispielsweise ablesen:

- negative und positive zufällige Fehler sind gleich wahrscheinlich;
- kleine zufällige Fehler sind wahrscheinlicher als große; sehr große zufällige Fehler sind extrem unwahrscheinlich;
- zwischen den Werten $\mu - \sigma$ und $\mu + \sigma$ (also -1 bis $+1$) liegen 68,3 % aller Fehler (die gestrichelte Fläche ist 68,3 % der Gesamtfläche unter der Kurve, die $= 1$ gesetzt ist).

Die abgebildete Kurve ist der Verlauf des Differentialquotienten der Verteilungsfunktion der *Normalverteilung* und heißt *Verteilungsdichte*.

Mit der beschriebenen Kurve könnte man arbeiten, wenn $n \to \infty$; (i. allg. kann diese Voraussetzung für $n \approx 200$ als erfüllt gelten). In der Praxis ist n jedoch üblicherweise wesentlich kleiner, d. h., man hat es statistisch gesehen nicht mit einer Grundgesamtheit, sondern mit einer Stichprobe zu tun. Die Standardabweichung der Stichprobe wird mit s (statt mit σ) bezeichnet und hat den Wert

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}\,.$$

Hierin ist $\bar{x}$ der arithmetische Mittelwert der Meßreihe: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, der sich umso weniger vom Erwartungswert μ unterscheidet, je größer n ist. Die in 7.2.6.3. bereits benutzte Standardabweichung s wird auch mittlerer quadratischer Fehler der Einzelmessung genannt und ist ein Maß für die Streuung der Meßreihe.

Wenn schon jedes x_i im Mittel mit dem quadratischen Fehler s behaftet ist, so muß der mittlere Fehler in $\bar{x}$ natürlich kleiner sein, und zwar wiederum umso kleiner, je größer n wird. Für diesen mittleren quadratischen Fehler $s_{\bar{x}}$ des Mittelwertes gilt $s_{\bar{x}} = s/\sqrt{n}$. Da s die Standardabweichung einer Stichprobe ist, hat $s_{\bar{x}}$ die gleiche statistische Sicherheit wie σ, also 68,3 %. Um auch die Streuung der Meßreihe für andere statistische Sicherheiten angeben zu können und ferner zu berücksichtigen, daß diese bei wenigen Meßwerten ($n \ll 200$) nur erreicht wird, wenn die sich ergebende Streuspanne vergrößert wird, muß $s_{\bar{x}}$ noch mit einem Faktor t (*Studentfaktor*) multipliziert werden. Tab. 7.2.6–6 enthält auszugsweise einige t-Faktoren. Wie man sieht, geht für $n \to \infty$ und $P = 68{,}3\,\%$ erwartungsgemäß der Studentfaktor $t \to 1$. Den durch Multiplikation mit t erhaltenen Wert $u_t = ts/\sqrt{n}$ nennt man *Vertrauensgrenze*, und der Bereich von

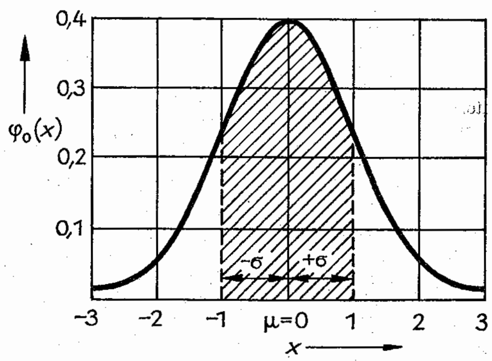


Abb. 7.2.6–5 Verteilungsdichte der standardisierten Normalverteilung

$\bar{x} - u_f$ bis $\bar{x} + u_f$ ist der Vertrauensbereich des aus der Meßreihe gewonnenen Mittelwertes. Damit ist auch für das Resultat einer Meßreihe das vollständige Meßergebnis anzugeben, und zwar

$[\bar{x} + (-e_s)] \pm (u_f + e_{sR})$ bzw. $\bar{x}_{korr} \pm u$.

u kann auch durch pythagoreische Addition aus u_f und e_{sR} gebildet werden (vgl. 7.2.6.3.). Natürlich muß für e_{sR} die gleiche statistische Sicherheit P gelten wie für u_f. Da e_{sR} ein geschätzter Wert ist, für den es oft nur sehr unsichere Anhaltspunkte gibt, spielt der Unterschied zwischen $P = 95\,\%$ und $P = 99\,\%$ keine Rolle; für $P = 68{,}3\,\%$ kann e_{sR} etwas knapper geschätzt werden.

Tab. 7.2.6–6 t-Faktoren für verschiedene statistische Sicherheiten

n	statistische Sicherheit 68,3 %	95 %	99 %
5	1,15	2,776	4,604
7	1,09	2,447	3,707
10	1,058	2,262	3,250
15	1,037	2,145	2,977
20	1,026	2,093	2,861
25	1,020	2,064	2,797
40	1,012	2,021	2,704
60	1,008	2,000	2,660
120	1,004	1,980	2,617

7.2.6.5. Meßunsicherheiten indirekter Messungen. Als Beispiel für eine indirekte Messung sei angenommen, daß sich die gesuchte Größe M aus den drei gemessenen Größen X, Y und Z berechnet, also $M = f(X, Y, Z)$. Die Meßgrößen seien mit den Unsicherheiten u_x, u_y und u_z gemessen und hinsichtlich der systematischen Fehleranteile korrigiert.

Dann beträgt die Unsicherheit in M:

$$u_M = \sqrt{\left(\frac{\partial M}{\partial X} u_x\right)^2 + \left(\frac{\partial M}{\partial Y} u_y\right)^2 + \left(\frac{\partial M}{\partial Z} u_z\right)^2}.$$

Allgemein gilt, auch für mehr als drei Meßgrößen, die in M eingehen, $u_M = \sqrt{\sum_{j=1}^{r} \left(\frac{\partial M}{\partial x_j} u_{xj}\right)^2}$.

Durch die pythagoreische Addition ist u_M größer als der größte Einzelwert der u_j, aber kleiner als die algebraische Summe $\sum_{j=1}^{r} u_{xj}$. Außerdem haben alle Summanden unter der Wurzel gleiche Einheiten, nämlich die Einheit von M^2, da $[x_j] = [u_{xj}]$, so daß sich die Einheiten der entsprechenden Faktoren wegkürzen. Schließlich bekommen die einzelnen Summanden durch die partiellen Ableitungen auch die richtige Wertigkeit; ist z. B. $M = f(Y^2)$, so erhält der Summand mit u_y den Faktor 2.

Nimmt man an, daß sich X, Y und Z als Mittelwerte aus Meßreihen ergeben, so lautet das vollständige Meßergebnis

$$M = f(\bar{x} + (-e_s)_x;\, \bar{y} + (-e_s)_y;\, \bar{z} + (-e_s)_z) \pm u_M.$$

Neben dem Fall der Berechnung einer gesuchten Größe M aus mehreren gemessenen (und damit fehlerbehafteten) Größen gibt es noch andere Möglichkeiten indirekter Messungen. Als Beispiel sei hier der Fall angeführt, daß die gesuchte Größe der Anstieg K_y einer Geraden $y = K_y x + K_0$ ist. Hier werden die y_i-Werte zu verschiedenen vorgegebenen x_i-Werten gemessen, als Meßresultate also n Wertepaare (x_i, y_i) erhalten. Werden diese Wertepaare in einem Koordinatensystem aufgetragen, so ergeben sich n Meßpunkte, die aufgrund der zufälligen Fehleranteile um die erwartete Gerade streuen.

Die Ermittlung der besten Geraden durch diese n Meßpunkte und damit auch des Anstieges K_y ist Aufgabe der *Ausgleichsrechnung*. Hierbei werden die x_i als fehlerfrei angenommen, so daß die gemessenen y_i mit positiven und negativen senkrechten Abweichungen $e_{fi} = y_i - K_y x_i - K_0$ um die Ausgleichsgerade streuen. Nach der Gaußschen *Methode der kleinsten Quadrate*, nach der auch die Formel für s abgeleitet wird, bringt man die Summe der Fehlerquadrate $\sum_{i=1}^{n} e_{f_i}$ zu einem Minimum. Aus den erhaltenen Gleichungen lassen sich K_y, K_0 und ihre Standardabweichungen s_{K_y} und s_{K_0} berechnen.

Die Ausgleichsrechnung läßt sich auch bei anderen funktionellen Zusammenhängen $y = f(x)$ anwenden. Es können damit die in diesen Zusammenhängen auftretenden Konstanten und ihre Standardabweichungen ermittelt werden. Voraussetzung ist lediglich, daß der Typ des funktionellen Zusammenhanges, der durch die Meßpunkte (x_i, y_i) meßtechnisch als Kurve ermittelt werden soll, aus theoretischen Überlegungen bekannt ist, die Funktion $y = f(x)$ also explizit angegeben werden kann.

7.2.7. Meßwertausgabe

7.2.7.1. Ausgabe analoger Meßwerte. Analoge Meßwerte werden normalerweise optisch mit Hilfe einer *Skale* und einer *Anzeigemarke* ausgegeben (vgl. 7.2.3.3. und 7.2.4.2.). Die Anzeigemarke muß nicht unbedingt ein Zeiger sein.

Nichtoptische Meßwertausgaben sind ausgesprochen selten. Gelegentlich kommen noch akustische Verfahren vor, z. B. beim Abgleich von Wechselstrombrücken mit Hilfe von Kopfhörern. In Spezialfällen gibt es auch eine taktile Ausgabe von Meßwerten, z. B. Blindenuhren.

Da analoge Meßgeräte ohne Speicher im Augenblick der Messung abgelesen werden müssen, eignen sich anzeigende Geräte vorzugsweise für die Messung zeitlich konstanter Größen. Bei zeitabhängigen Größen hilft man sich durch Stoppen des Meßgerätes, z. B. Stoppuhr, oder durch das Festhalten markanter Meßwerte, z. B. Extrema. Maximum-Minimum-Thermometer halten beispielsweise die höchste und niedrigste Temperatur im Zeitraum zwischen zwei Ablese- und Rückstellvorgängen fest. Den gleichen Effekt erreicht man bei Zeigermeßgeräten durch sogenannte Schleppzeiger, die vom Zeiger des Meßwerkes mitgenommen werden aber nicht von allein mit zurückgehen. Voraussetzung ist dabei, daß ausreichende Meßkräfte zur Verfügung stehen.

Es ist weiterhin möglich, den Anzeigegeräten bestimmte Operationen zu übertragen, z. B. das Auslösen von Alarmsignalen oder die Betätigung von Schaltern. Ein *Quecksilber-Kontaktthermometer* z. B., wie es in Thermostaten benutzt wird, kann beim Erreichen der Schalttemperatur über den Quecksilberfaden einen Stromkreis schließen, der mit Hilfe eines Relais die Thermostatenheizung abschaltet. Bei Zeigermeßgeräten können Schaltvorgänge über induktive, optoelektronische u. a. Abgriffe ausgelöst werden. Stehen große Meßkräfte zur Verfügung, z. B. bei Ausdehnungs- und

Bimetallthermometern oder Schwimmerfüllstandsmeßgeräten, so können auch mechanische Schalter direkt betätigt werden.

Der Dokumentation analoger Meßwerte dienen *Registriergeräte*. Es werden die unterschiedlichsten Aufzeichnungsverfahren, mechanische, elektrische, fotografische u. a., benutzt. Während ein Fallbügelschreiber ein Punktschreiber mit diskontinuierlichem Ausgangssignal ist, wird von einem *Kompensationsbandschreiber* ein kontinuierliches Ausgangssignal abgegeben (Linienschreiber). In einem Motorkompensator, dessen Grundprinzip Abb. 7.2.4–1b zeigt, wird dem Verstellmechanismus für die Kompensationsspannung der Schreibmechanismus parallel geschaltet. Da die Kräfte für das Schreiben (zur Überwindung von mechanischen Trägheiten, von Reibungswiderständen des Schreiborgans auf dem Papier usw.) vom Motor aufgebracht werden, bleibt die Rückwirkungsfreiheit des Kompensationssystems erhalten. Es sind also hohe Genauigkeiten erreichbar.

7.2.7.2. Digitale Meßwertausgabe. Im Fall der digitalen Ausgabe von Meßwerten ist der Zahlenwert direkt in Ziffernform abzulesen, und zwar als Dezimalzahl. Bei mechanischen und elektromechanischen Zählern sind die der Anzeige dienenden Ziffern direkt auf den Zählrollen untergebracht. Als Beispiel kann ein einfacher Kilometerzähler für Fahrräder (Abb. 7.2.7–1) dienen, der durch ein Transportzahnrad und einen am Vorderrad befestigten Mitnehmer betätigt wird. Andere mechanische Digitalausgaben lassen sich durch Zifferntafeln, die elektromechanisch „umgeblättert“ werden, realisieren, wie sie von der Angabe der Abflugzeiten auf Flughäfen bekannt sind. Als weitere Anzeige, die auch aus größerer Entfernung erkennbar ist (*Großsichtanzeigen*), kommt die Projektion nach dem Prinzip des Bildwerfers in Frage. Eine andere Möglichkeit ist die *Mosaikanzeige*, die meist aus 35 Lichtpunkten besteht. Für die Ansteuerung solcher Anzeigen sind aufwendige Codeumsetzer erforderlich.

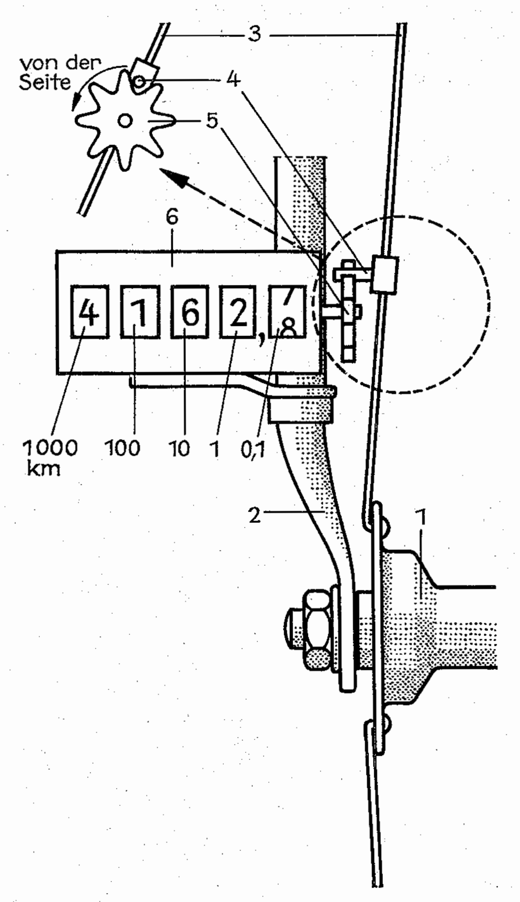


Abb. 7.2.7–1 Vereinfachte Darstellung eines Kilometerzählers für Fahrräder

Die weiteste Verbreitung für die digitale Meßwertausgabe haben in den letzten Jahren die *Segmentziffernanzeigen* gefunden, obwohl bei den vielfach benutzten sieben Segmenten die Ziffern in ihrer Darstellung sehr stark vereinfacht sind. Durch die Verbreitung der Digitaluhren und elektronischen Taschenrechner ist diese Form der Zifferndarstellung allgemein vertraut. Die einzelnen Segmente sind entweder Lumineszenzdioden oder Flüssigkristalle (LCD-Bauelemente [engl. *l*iquid *c*rystal *d*isplay]), wobei die Bedeutung der letzteren ständig zunimmt. Während Lumineszenzdioden selbstleuchtend sind, benötigen Flüssigkristallanzeigen Fremdlicht. Dafür ist ihre Leistungsaufnahme ($<0{,}1\ \mathrm{mW \cdot cm^{-2}}$) erheblich geringer. Da die Codeumsetzer für die Ansteuerung von 7-Segment-Anzeigen als integrierte Schaltkreise verfügbar sind, können derartige Anzeigeeinheiten nahezu beliebig klein gehalten werden.

Für die Registrierung und Speicherung digitaler Meßwerte steht eine hoch entwickelte Technik zur Verfügung, denn man kann sich dabei der gleichen Verfahren bedienen, die in der elektronischen Rechentechnik für die Datenspeicherung und -ausgabe benutzt werden. In der Meßtechnik spielen vor allem mechanische und magnetische Verfahren eine Rolle. Zu den mechanischen Verfahren gehören *Drucker* und *Stanzer*.

7.2.8. Automatisierte Meßwertverarbeitung

7.2.8.1. Bedeutung der Meßwertverarbeitung. Im weitesten Sinne kann man unter der automatisierten Meßwertverarbeitung jede Verwendung von Meßwerten verstehen, die über die in 7.2.7. behandelte Meßwertausgabe hinausgeht. Eine automatische Meßwertverarbeitung liegt also vor, wenn ohne menschliches Zutun

- aus mehreren Meßwerten ein Meßergebnis, z. B. bei Mittelwertbildung und indirekter Messung, errechnet wird,
- rechnerische Korrekturen an Meßwerten angebracht werden,
- Meßwerte nach vorgegebenen Kriterien sortiert oder ausgesondert werden,
- Meßwerte für Schaltvorgänge o. a. Operationen genutzt werden,
- Meßwerte als Eingangs- oder Führungsgrößen für Regler dienen,
- Meßwerte als Grundlage für das Sortieren von Objekten verwendet werden, z. B. in der Qualitätsüberwachung, sowie
- Meßwerte in Berechnungen, Entscheidungsalgorithmen u. a. logische Operationen eingehen usw.

Die meisten dieser Aufgaben werden oft als eigenständige

Gebiete betrachtet, so daß zur automatischen Meßwertverarbeitung im engeren Sinne schwerpunktmäßig solche Aufgaben gerechnet werden, wie sie im letzten Punkt der vorstehenden Aufzählung angedeutet sind. Mit dem Vordringen der Mikroprozessoren und der Mikrorechentechnik lassen sich solche Aufgaben mehr und mehr dezentral lösen.
In den folgenden Abschnitten werden einige Beispiele für automatische Rechenoperationen in Verbindung mit Meßgeräten betrachtet, ohne einzelne Aufgabengebiete genauer zu unterscheiden. Gerade die Entwicklung auf diesem Gebiet ist so stark, daß sich ständig neue Möglichkeiten eröffnen und kaum abzusehen ist, wie weit die Mikroelektronik noch in das Gesamtgebiet der Meßtechnik eindringen und diese zum Teil grundlegend verändern wird.
7.2.8.2. Analoge Meßwertverarbeitung. Die Meßwertverarbeitung unter Verwendung analoger Signale hat gegenüber der Digitaltechnik an Bedeutung verloren; sie wird deswegen aber nicht gänzlich überflüssig werden. Komplette *Analogrechner* werden allerdings nur noch in seltenen Fällen für Meßwertverarbeitungszwecke eingesetzt; man benötigt sie höchstens, wenn einzelne Produktionsabschnitte oder Regelkreise einschließlich der zugehörigen Meßtechnik modelliert werden, um beispielsweise ihr Zeitverhalten zu untersuchen.
Einzelne analoge Rechenglieder oder -schaltungen findet man dagegen noch häufiger. Sie dienen einmal dazu, einfache Korrekturen durchzuführen, z.B. Kennlinien zu linearisieren, Temperaturfehler zu korrigieren oder Geräte hinsichtlich ihres Zeitverhaltens zu verbessern. Das zweite Aufgabengebiet ist die *Verknüpfung* mehrerer *gemessener Werte*. Abb. 7.2.8-1 zeigt beispielsweise eine Meßeinrichtung zur Bestimmung des Massenstromes $\dot{m}$ einer Flüssigkeit. Die gemessenen Werte für den Volumenstrom $\dot{V}$ und die Flüssigkeitsdichte ϱ werden in elektrische Einheitssignale umgewandelt und diese in einer einfachen analogen Rechenschaltung multipliziert, denn $\dot{m} = \dot{V}\varrho$. Sind die Einheitssignale Stromstärken, so kann zur Produktbildung ein elektrodynamisches Meßwerk benutzt werden, das direkt einen dem Produkt der Ströme proportionalen Ausschlag hat, der in Einheiten von $\dot{m}$ graduiert werden kann. Benötigt man ein elektrisches Ausgangssignal, so kann man einen Hallgenerator zur Produktbildung verwenden. Für genauere Rechenoperationen trifft man auch noch elektromechanische Momentvergleichsrechner an.
Neben den elektrischen gibt es vereinzelt auch noch pneumatische oder rein mechanische Verfahren. Zu ihnen gehören pneumatische Momentvergleichsrechner und Kurvenscheiben zur Kennlinienlinearisierung. Derartige Verfahren werden zunehmend durch elektronische abgelöst, die mit digitalen Signalen arbeiten.
7.2.8.3. Digitale Meßwertverarbeitung. Durch die überaus stürmische Entwicklung der Mikroelektronik und der maschinellen Rechentechnik hat die digitale Meßwertverarbeitung in den letzten Jahren einen ungeahnten Aufschwung genommen. Dabei lassen sich einerseits Rechner unterschiedlicher Größenordnung, die für die Lösung von vielerlei Problemen benutzt werden können, und andererseits einzelne Rechenglieder oder Kleinstrechner mit spezialisierten Aufgabenstellungen unterscheiden, die nur wenige festprogrammierte Operationen auszuführen haben. Voraussetzung ist in jedem Fall, daß die zu verarbeitenden Signale in digitaler Form vorliegen. Analoge Signale müssen vor der Verarbeitung digitalisiert werden, wobei für einfache Rechenoperationen zu prüfen ist, ob nicht eine analoge

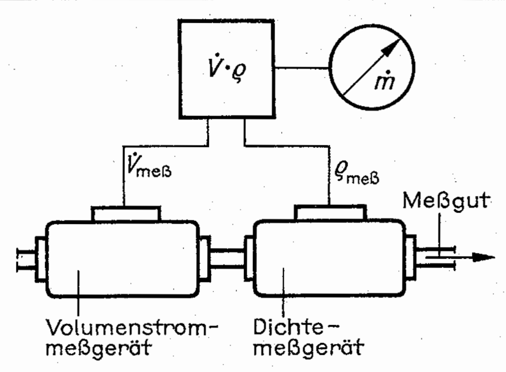


Abb. 7.2.8-1 Multiplikationsglied zur Ermittlung des Massestromes $\dot{m}$ als Beispiel für eine analoge Rechenschaltung

Verarbeitung mit entsprechend geringerem Aufwand für den beabsichtigten Zweck ausreicht.
Mikroprozessoren und Kleinstrechner werden jetzt zunehmend in Meßgeräte integriert (vgl. 7.2.8.4.). Dabei überträgt man ihnen die unterschiedlichsten Aufgaben, nicht nur solche der eigentlichen Meßwertverarbeitung.
Eine andere Situation ergibt sich, wenn mehrere verschiedene Meßgeräte mit einer zentralisierten *Meßwertverarbeitungsanlage* (MWVA) zusammenarbeiten. In diesem Fall wird auch nicht gefordert, daß die Meßgeräte digitale Ausgangssignale liefern, da die MWVA einen zentralen AD-Umsetzer aufweist. Neben analogen und digitalen Meßsignalen können zusätzliche Daten, z. B. Sollwerte, eingegeben werden. Die einzelnen Meßstellen können je nach Programm in beliebiger Folge und Häufigkeit abgefragt werden; im Bedarfsfall werden die Werte vorübergehend in einem Puffer zwischengespeichert. Andere Speicher enthalten Programme für den Funktionsablauf und die Rechenoperationen. Die verarbeiteten Informationen können in verschiedener Form analog und digital ausgegeben werden. Zentralisierte MWVA werden vor allem zur Prozeßanalyse, insbesondere in Vorbereitung eines Prozeßrechnereinsatzes oder für Zwecke der Betriebsführung verwendet, wobei es vor allem um eine nach unterschiedlichen Gesichtspunkten vorzunehmende Datenverdichtung geht. Der Aufwand für eine zentralisierte MWVA ist recht erheblich, nicht zuletzt wegen der erforderlichen Leitungen für eine störungsfreie Übertragung der Meßdaten.
Während die Bedeutung derartiger MWVA sicher weiter zurückgeht, bleibt die Meßwertverarbeitung durch zentrale Rechner, die nicht speziell für diese Aufgabe vorgesehen sind, sondern vorzugsweise andere Aufgaben haben, auch zukünftig von Interesse. Hierbei sind die sog. peripheren Geräte, das sind in diesem Fall die Meßgeräte, Meßstellenumschalter, Bedien- und Ausgabeeinheiten usw. über ein Interface mit dem Rechner verbunden. Dieser wird nur durch entsprechende Programmierung der jeweiligen Meßwertverarbeitungsaufgabe angepaßt. Es werden also keine Spezialrechner benötigt.
Ein *Interface* ist die nach vereinbarten konstruktiven, elektrischen und logischen Bedingungen gestaltete Schnittstelle zwischen Rechner und Peripherie mit den entsprechenden Baugruppen im Ein-/Ausgabe-Kanal des Rechners und im Steuerteil von Geräten und Funktionseinheiten. Um möglichst unterschiedliche Geräte ohne Komplikation miteinander verbinden zu können, sind die Interfaces standardi-

siert. Allerdings gibt es aus mehreren Gründen eine ganze Reihe von Standard-Interfaces (SI). Für die Meßtechnik von Bedeutung sind vor allem SI 1.2., CAMAC und das byteserielle-bitparallele Interface, IEC-Bus genannt.

7.2.8.4. Intelligente Meßgeräte. Werden formalisierbare geistige Tätigkeiten einzelnen Klein- und Kleinstrechnern übertragen statt einem größeren zentralen Rechner, so spricht man von verteilter Intelligenz. Dementsprechend bezeichnet man auch Meßgeräte, die einen Mikroprozessor oder Kleinrechner für die Ausführung mehr oder minder komplizierter geistiger Tätigkeiten enthalten, die früher dem Messenden oblagen, als *intelligente Meßgeräte*. Die Struktur, der Aufbau und die den Recheneinheiten übertragenen Aufgaben unterscheiden sich von Meßgerät zu Meßgerät sehr stark, so daß keine allgemein verbindliche Darstellung eines intelligenten Meßgerätes gegeben werden kann. Außerdem sind die von den Prozessoren übernommenen Funktionen nicht nur Meßwertverarbeitungsprobleme, sondern liegen auch in anderen Bereichen. Einige Beispiele sollen das illustrieren.

Erhöhung des Bedienkomforts. Bei Meßanordnungen, die für die Durchführung einer Messung mehrere verschiedene Operationen erfordern, wie Probenzuführung, Justieren, Messen, Probenwechsel, läßt sich der Ablauf voll automatisieren. Auch notwendige Einstellvorgänge, wie Wahl des günstigen Meßbereiches, Beachtung der Polung einer Gleichspannung, Vorsorge gegen Überlastung des Meßwerkes u. a. kann dem Mikroprozessor übertragen werden.

Erhöhung der Meßgenauigkeit ist ein besonderer Vorteil intelligenter Meßgeräte. Dazu kann dem eingebauten Rechner die Anbringung einer Korrektion entsprechend der Fehlerkurve (Beseitigung systematischer Fehleranteile, vgl. 7.2.6.2.) übertragen werden. Außerdem kann durch Korrektur die Auswirkung von Störgrößen berücksichtigt werden, die durch eine getrennte Messung laufend erfaßt und dem Meßgerät zugeführt werden. Schließlich ist es auch möglich, vom Rechner das Gerät in regelmäßigen Zeitabständen abgleichen zu lassen, um zeitabhängige Fehlerursachen zu erkennen und auszuschalten. Darüber hinaus kann das dynamische Verhalten des Gerätes verbessert, also der Einfluß dynamischer Fehler verkleinert werden.

Meßdatenverdichtung. Aus aufgenommenen Meßreihen kann der Rechner Mittelwert und Streuung (Standardabweichung) sowie andere statistische Kenngrößen ermitteln. Diese Auswertung kann sogar dahingehend modifiziert werden, daß der eingebaute Rechner z.B. anhand der Standardabweichung entscheidet, ob der Umfang der Meßreihe zu vergrößern oder zu verkleinern ist. Weiterhin können mehrere Meßwerte rechnerisch zu einem Meßergebnis zusammengefaßt werden (indirekte Messungen). Auch die Multiplikation mit einem konstanten oder variablen Faktor ist möglich, z. B. kann bei Mengenmeßwerten der zugehörige Preis laufend mit ausgegeben werden.

Meßdatenbeurteilung. Ein eingebauter Rechner kann die erhaltenen und korrigierten Meßwerte auch gemäß vorgegebener Kriterien beurteilen. Er kann offensichtlich falsche und mit groben Fehlern behaftete Werte aussortieren oder die Einhaltung bestimmter Toleranzgrenzen kontrollieren und bei Überschreitung Alarm auslösen. Auch eine Klassierung der Meßwerte ist möglich, und sogar die Meßwertverteilung kann daraus ermittelt werden. Durch eine Analyse der Änderungsgeschwindigkeit der Meßgröße kann der Rechner entscheiden, ob die Häufigkeit der Messungen bei einer zeitabhängigen Größe ausreichend ist oder nicht.

Mit diesen Beispielen sind die Möglichkeiten, die intelligente Meßgeräte bieten, noch keineswegs erschöpft. Sie haben aber gezeigt, daß die Mikroelektronik die Meßtechnik grundlegend revolutionieren wird. Damit ist aber nicht gesagt, daß die einfachen anzeigenden Meßgeräte nicht auch künftig für eine große Zahl von Anwendungsfällen ihre Bedeutung behalten werden.

7.3. Meßdatengewinnung

7.3.1. Messung von Längen, Winkeln, Flächen, Volumina

7.3.1.1. Allgemeines zur Längenmessung. Das Meter ist die Länge der Strecke, die Licht im Vakuum während der Dauer von 1/299 792 458 Sekunden durchläuft. Diese Definition basiert auf einem festgelegten Wert für die Lichtgeschwindigkeit c im Vakuum. Für die praktische Handhabung wird die oben definierte Längeneinheit auf körperliche Vergleichsmaßstäbe übertragen, die Abstandsmarken (für bestimmte Temperatur und weitere fixierte Umgebungsbedingungen) tragen. Die Genauigkeit solcher Vergleichsmaßstäbe beträgt etwa einige 10^{-7}, d. h., für $l = 1$ m beträgt die Unsicherheit (der prinzipiell nicht unterschreitbare Fehler) einige 10^{-7} m in der Längenangabe. Die Übertragung der gegebenen Längeneinheit auf einen solchen Vergleichsmaßstab ist mit einem Fehler von etwa 10^{-7} behaftet (ist also drei bis fünfmal kleiner als der Fehler des Maßstabes selbst). Die Anschlußmessung an eine geeignete Strahlungsquelle (z. B. Laser), sowie die Kontrolle solcher Maßstäbe erfolgt mittels Interferometer (↗ Optik). Die Ablesung von Abstandsmarken erfolgt unter Meßmikroskopen mit einem Komparator. Als sehr genaue praktische Meßmittel für absolute und relative Längenmessungen zwischen ≈ 0,1 mm und allgemein 0,25 m (aber auch bis max. 2 m) gelten *Endmaße*. Das Längenmaß ist dabei durch den Abstand zweier paralleler Flächen definiert. Durch spezielle metallurgische Behandlung wird eine hohe Konstanz erreicht. Die Endflächen sind gehärtet und geschliffen (optische Qualität). Infolge der hochglanzpolierten Endflächen lassen sich mehrere solcher Endmaße „anschieben“, wobei

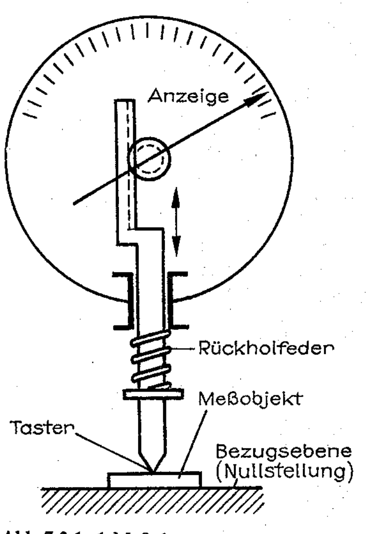


Abb. 7.3.1–1 Meßuhr

hohe Anziehungskräfte zwischen den Flächen wirksam werden, so daß mehrere solcher Stücke freitragend aneinander haften. Die Genauigkeit von Endmaßen liegt bei einigen Zehntel Mikrometer. Innerhalb von mehreren Jahren können aber Abweichungen bis zu 2 µm auf 100 mm Länge eintreten. Die Endmaße sind deshalb in Abständen einiger Jahre neu zu bestimmen. Wegen der hohen optischen Qualität ihrer Endflächen können sie direkt mit einem Interferometer vermessen werden. Mit einem Normsatz Endmaße lassen sich alle Maße zwischen 1,000 mm und mehr als 200 mm in Abstufungen von 0,001 mm herstellen. Endmaße verlangen eine sehr sorgfältige Behandlung; sie gehören nur selten zu den alltäglichen Meßmitteln.
Häufiger anzutreffen sind *Meßschraube, Schiebelehre* und *Strichmaße* sowie *Lehren für feste Maße*. Mittels einer *Meßschraube* lassen sich Längen zwischen 0,01 mm und meist 25 mm (auch bis 100 mm und mehr) auf etwa 0,005 mm genau messen. Eine Meßschraube enthält eine Längsteilung von 0,5 mm und die Schraube (übliche Steigung 0,5 mm) trägt auf dem Umfang eine meist 50teilige Skale. Außer der Meßschraube verwendet man noch Meßuhren oder Fühlhebel (Abb. 7.3.1–1). Die Genauigkeit kann dabei durch mechanische Hebelübersetzungen oder/und optische Mittel gesteigert werden. Die erreichbare Genauigkeit liegt bei ≈1 µm, in Ausnahmefällen 0,2 µm.
Bei der *Schublehre* erfolgt die Ablesung der Länge auf dem Maßstab mittels Nonius. In den beiden letzten Fällen macht man vom Komparatorprinzip Gebrauch.
Strichmaße bestehen meist aus Metall (Stahlmaß), für optische Geräte auch aus Glas oder Quarz; für geringe Genauigkeiten auch aus Plaste oder Holz. Bandmaße verschiedenster Genauigkeiten sind bis zu 50 m Länge (in Einzelfällen auch mehr) in Gebrauch. Strichmaße tragen für die freie Ablesung eine mm-Teilung mit Strichdicken zwischen 0,05 und 0,2 mm. Bei Ablesung oder Registrierung mittels optischer Mittel (auch optoelektronischer Meßfühler in Automaten) werden Teilungen bis 1/100 mm mit Strichdicken zwischen etwa 1 und 3 µm hergestellt. Sie erfordern zur Ablesung eine 20- bis 50fache Vergrößerung, z. B. Meßmikroskop.
In der Feinwerktechnik werden in zunehmendem Maße Universallängenmesser und Winkelmeß- und Teilungs-Einrichtungen, meist mit digitaler Registrierung, auch zur Steuerung von Bearbeitungsautomaten eingesetzt. Die Genauigkeiten liegen bei einigen Zehntel Mikrometern bzw. Winkelsekunden.

7.3.1.2. Winkelmessung. Die Winkelteilung wird am Umfang einer Kreisskale abgelesen oder registriert, d. h. auf eine Längenmessung zurückgeführt, weil der Kreisbogen dem Winkel proportional ist. Als genaueste Registrierung eines Winkels innerhalb eines kleinen Bereiches (Feintrieb) bezogen auf eine feste Marke (oder Anschlag) gilt gegenwärtig $3 \cdot 10^{-8}$ rad/µm (in einem Bereich von 10^{-4} rad). Große Spiegelteleskope (Masse der Halterung größenordnungsmäßig 100 Tonnen) sind auf mindestens 1″ genau einstellbar. In der Astronomie lassen sich (z. T. mittels fotografischer Aufnahmen) noch Winkel unter 0,01″ messen. Mit interferometrischen Methoden sind noch bis zu einige 100mal kleinere Winkel bestimmbar. Die zur Winkelmessung verwendeten Instrumente sind 7.3.1.3. und 4 erwähnt.

7.3.1.3. Mechanische Längenmeßverfahren. Längenmessung wird in höchster Präzision beim *Komparator* verwirklicht. Dieser wird zum Vergleich zweier Maßstäbe (Präzisionsstrichmaße) und vor allem zum Auswerten von

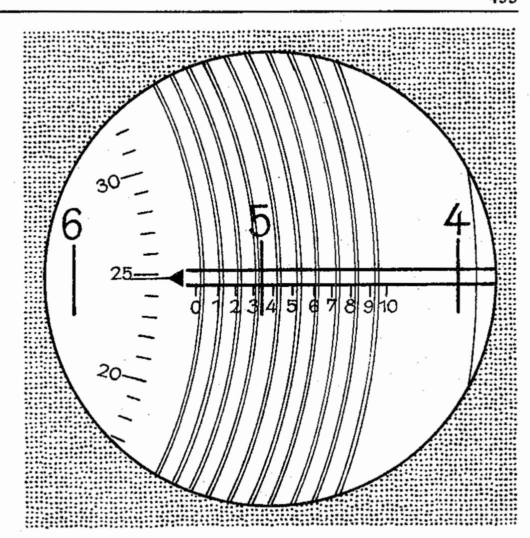


Abb. 7.3.1–2 Sehfeld des Spiralmikroskopes zur Ablesung am Komparator; 1 Umdrehung der Spirale ≙ 0,1 mm, Ablesung: 5,324(8) mm

Spektralplatten (fotografische Aufnahmen) verwendet. Die Unsicherheiten beim Vergleich zweier Strichmarken liegen im besten Fall bei 0,2 µm. Hierbei wird der Prüfling verschoben. Die Abb. 7.3.1–2 zeigt die Ablesung am Komparator. Bei den Messungen muß eine längere Zeit vor der Messung und während dieser die Temperatur auf wenigstens 1/10 Grad konstant gehalten werden.
Für Längenmessungen mit dem Mikroskop ist ein *Meßschraubenokular* sehr zweckmäßig. Innerhalb des Okulars ist eine Strichplatte mittels Mikrometertriebes verschiebbar (und außen ablesbar). Damit kann man das Objekt genau vermessen. Anschließend ersetzt man das Objekt durch eine Strichplatte bekannter Teilung und hat somit unab-

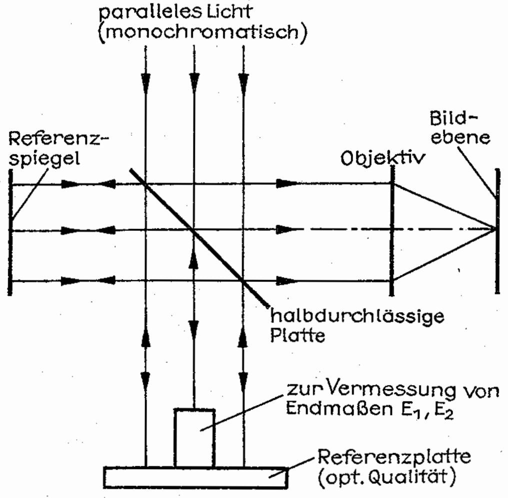


Abb. 7.3.1–3 Prinzip des Interferenzkomparators (zur Vermessung von Endmaßen E_1, E_2)

hängig vom Abbildungsmaßstab der Objektive und einer eventuellen Änderung der Tubuslänge (z. B. durch Analysatoren bei polarisiertem Licht) sofort eine Objektgrößenangabe.

7.3.1.4. Optische Längenbestimmung. Diese Methode besteht prinzipiell im Auszählen von Wellenlängen des zur Messung verwendeten Lichtes. Je kürzer die Wellenlänge ist, desto genauer kann man messen. Im allgemeinen sind diese Verfahren auf sichtbares Licht beschränkt ($\lambda \approx 0{,}4$ bis 0,7 µm). Die ausmeßbare Strecke ist außer von der Konstruktion der Apparatur grundsätzlich von der Kohärenzlänge der verwendeten Strahlung abhängig. Mit Lasern (große Kohärenzlänge) lassen sich auf diese Weise Endmaße von 1 m und darüber vermessen.

In der Praxis werden Endmaße mit *Interferenzkomparatoren* (Ausführung für relative und absolute Längenmessung) oder Interferenz-Doppelprismen vermessen. Das Prinzip ist aus Abb. 7.3.1–3 erkennbar und das Gesichtsfeld zeigt Abb. 7.3.1–4.

Dünne Drähte werden ebenfalls zweckmäßig interferometrisch vermessen. Nach Abb. 7.3.1–5 wird eine planparallele dicke Glasplatte 2 schräg auf den Draht (oder einen anderen Körper) über die Glasplatte 1 gelegt. Die Beleuchtung erfolgt mit monochromatischem Licht der Wellenlänge λ. Es wird die Anzahl n der Streifen beobachtet. Die Drahtdicke ist dann $d = n \cdot \lambda/2$ (n = Plattenlänge/Streifenabstand). *Dünne Schichten* mißt man auf ähnliche Weise.

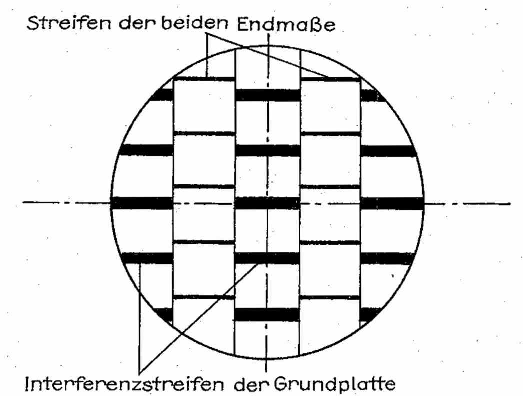


Abb. 7.3.1–4 Prinzipielles Bildfeld eines Interferenzkomparators mit zwei Endmaßen (monochromatisches Licht)

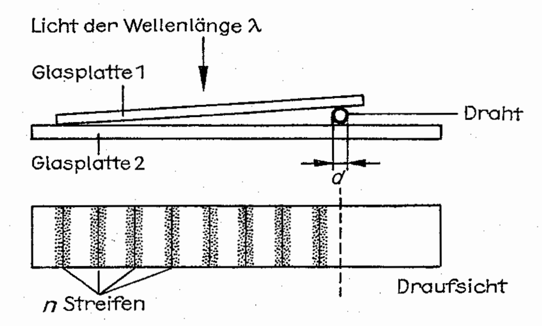


Abb. 7.3.1–5 Interferometrische Messung dünner Drähte, z. B. $\lambda = 600$ nm, $n = 8$, $d = \lambda/2 \cdot n = 2{,}4$ µm im Durchmesser

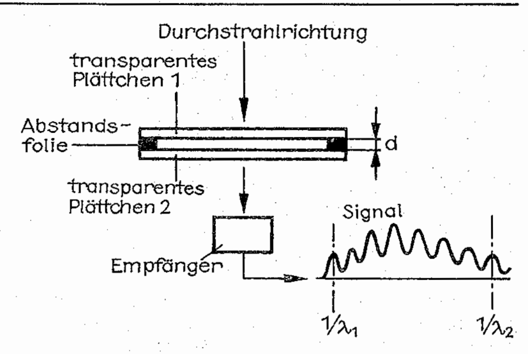


Abb. 7.3.1–6 Dickenmessung d einer Meßzelle mittels Interferenzen in einem Spektralfotometer; λ_1, λ_2 = Wellenlängen

Der Abstand zweier paralleler durchsichtiger Plättchen (z. B. einer Meßzelle) ergibt sich bei Verwendung eines Spektralfotometers zwischen den Wellenlängen λ_1 und λ_2 zu d = (Anzahl der Maxima)/$[2n(1/\lambda_1 - 1/\lambda_2)]$; für n (Brechungsindex) = 1 und Wellenlängen zwischen 0,4 und 0,7 µm lassen sich Dicken d zwischen 2 und 60 µm ausmessen (Abb. 7.3.1–6). Die Dicke dünner Schichten (aufgedampft, gesputtert o. ä.) bis herab zu weniger als 10^{-3} µm mißt man u. a. interferometrisch. Bis zu etwa 0,01 µm gibt es mechanisch-elektronische Abtastsysteme (auch berührungslos). Mit einem Interferenzmikroskop lassen sich noch Dickenänderung (geeigneter Objekte) bis auf weniger als 10^{-3} µm erkennen.

7.3.1.5. Trigonometrische Entfernungsmessung. In den Bereichen Landvermessung, Bauwesen, Bergbau, aber auch im Großmaschinenbau u. ä. werden weitgehend optische Entfernungs- und Winkelmeßeinrichtungen eingesetzt. Prinzipiell wird dabei aus genauen Winkelmessungen von einer bekannten Basis aus mit Hilfe der ebenen Trigonometrie die Entfernung zum Zielpunkt bestimmt. Die verwendeten Winkelmeßgeräte sind *Theodolite* (eventuell *Sextant*), *Nivelliere* bzw. *Tachymeter*. Theodolite bestehen im Prinzip aus einem Fernrohr (10 bis 60fache Vergrößerung), das in zwei exakt senkrecht aufeinanderstehenden Achsen (horizontal und vertikal) sehr genau einstellbar ist (Abb. 7.3.1–7). Die Winkelteilungen sind je nach Typ zwischen 0,1″ und 10″ ablesbar. Die Abweichung von der Horizontalen wird durch einen Kompensator (Schwerependel) auf kleiner als 10″ bis 1′ gehalten. In einfachsten Geräten verwendet man zur Anzeige bzw. zur Kontrolle der Horizontalen eine Libelle (durchsichtiges Gefäß), die bis auf ein kleines freies Volumen (Blase) mit Flüssigkeit gefüllt ist. Die *Dossenlibelle* ist am Deckel innen kugelförmig ausgeschliffen und trägt eine Kreismarke. Die Empfindlichkeit beträgt etwa 1′ bis 5′. Die *Röhrenlibelle* besteht aus einem Glaszylinder, der innen nach einem Kreisbogen vom Radius r (tonnenförmig) ausgeschliffen ist. Für $r \approx 20$ m erreicht man eine Empfindlichkeit von 20″/2 mm Markenabstand. Ein einfacher künstlicher Horizont besteht aus Quecksilber, dessen Oberfläche (einige Quadratzentimeter) sich in einem dosenförmigen Glasgefäß sehr gut horizontal einstellt (vielfach mit Sextant in Gebrauch).

Die derzeitig modernsten Theodoliten (Tachymeter) ermitteln nach Anvisieren des Zielpunktes, in dem sich jedoch ein Reflektor für die verwendete Strahlung befinden muß, automatisch mittels Optoelektronik und eingebautem

Kleinrechner die Entfernung (zwischen einigen Metern und einigen tausend Metern je nach Typ und Sichtverhältnissen mit einer relativen Genauigkeit von 10^{-5} bis 10^{-6}) und die Winkel (bezüglich einer Basis horizontal und vertikal auf 1″ genau). Die Anzeige erfolgt digital. Bei diesen Geräten wird die Laufzeit des Signals zwischen Gerät und Ziel (und zurück) aus der Phasendifferenz zwischen abgehendem und ankommendem Modulations-Signal ermittelt. Diese Modulation wird einem Lichtemitter, meist eine Laserdiode, aufgeprägt. Bei einer Modulationsfrequenz von z. B. 10^6 Hz entspricht eine Periode 10^{-6} s und somit einer Gesamtstrecke von 300 m. Da man noch sehr kleine Phasenwinkel messen kann, ergeben sich hohe Genauigkeiten. Man verwendet in der Praxis mehrere Frequenzen (z. B. 0,15; 1,5 und 15 MHz für die Bereiche 1 km, 100 m, 10 m). Die Entfernungsmessung mit bekannter Basis im Ziel (Nivellierlatte) beruht auf dem Vergleich des Winkels, unter dem eine bestimmte Abstandsmarke auf dieser Latte im Fernrohr erscheint, mit einer Meßmarke im Fernrohr, deren Winkel (entsprechend der Konstruktion) bekannt ist. Bei nicht zugänglichen Zielen verwendet man Schnitt- und/oder Doppelbild-Entfernungsmesser. Dabei erzeugt die verwendete Optik vom Ziel zwei Bilder (oder zwei verschobene Hälften), die durch richtungsändernde Spiegel, Prismen oder Meßkeile zur Deckung gebracht werden bzw. ein geschlossenes Bild ergeben. Die dazu erforderliche Verstellung der Optik ist ablesbar und liefert bis zu einigen 1000 m ein Maß für die Entfernung. Zur Entfernungsmessung im Verkehrswesen, insbesondere für größere Entfernungen, werden spezielle Radargeräte verwendet, z. B. Hafenradar, Blindlanderadar. Zur Strukturauflösung muß sich dabei aber das Ziel genügend gut bezüglich des Reflektionsverhaltens von der Umgebung unterscheiden. Die Entfernung ist proportional der Laufzeit eines impulsförmigen Signals, wobei die Impulsdauer sehr viel kleiner als die Laufzeit sein muß.

Zur Vermessung größerer Entfernungen auf der Erde benötigt man den Erdradius. Für einfache Zwecke reicht es aus, einen mittleren Erdradius zu definieren, der aus dem Volumen einer als kugelförmig gedachten Erde folgt. Große Entfernungen auf der Erde vermißt man mit Hilfe der sphärischen Trigonometrie und unter Zuhilfenahme der Gestirne (geografische Breite) und genauer Zeitvergleiche (geografische Länge).

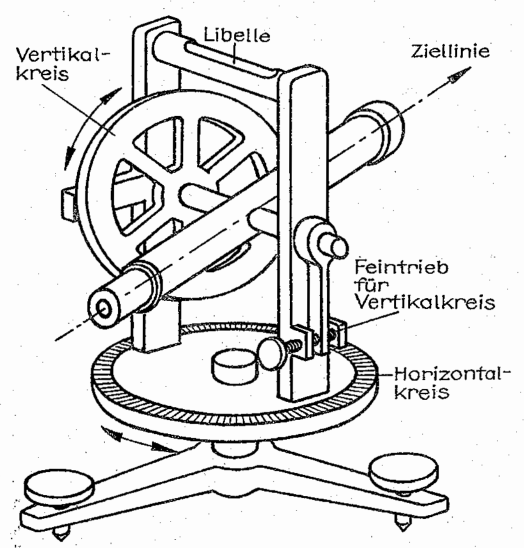


Abb. 7.3.1–7 Prinzip eines Theodoliten

Entfernungen im Sonnensystem folgen als Verhältniszahlen aus dem 3. Kepler'schen Gesetz $a^3/T^2 = \text{const}$. Es ist daher erforderlich, wenigstens eine große Halbachse a einer Planeten- oder Planetoidenbahn absolut zu bestimmen. In der zweiten Hälfte des 17. Jahrhunderts wurde erstmalig die Marsparallaxe im Periheldurchgang gemessen. Als Basis diente die Entfernung Cayenne (Mittelamerika) – Paris (exakt die Sehne des zugehörigen Großkreises).

Als Marsparallaxe ergaben sich etwa 24″. Das ist der Winkel, unter dem vom Mars aus die vermessene Basis a auf der Erde erscheint, d. h. die betreffende Entfernung betrug $5{,}69 \cdot 10^7$ km. Die weitere Umrechnung führte auf eine Sonnenparallaxe von etwa 9″, d. h. eine mittlere Entfernung Erde–Sonne von etwa $1{,}5 \cdot 10^8$ km. Der heute genaueste Wert beträgt 8,794 18″ (Äquatorradius der Erde von 6 378,33 km). Die genauesten Werte lassen sich aus Radarbeobachtungen gewinnen.

Trigonometrische Parallaxe von Sternen. Für Entfernungs-

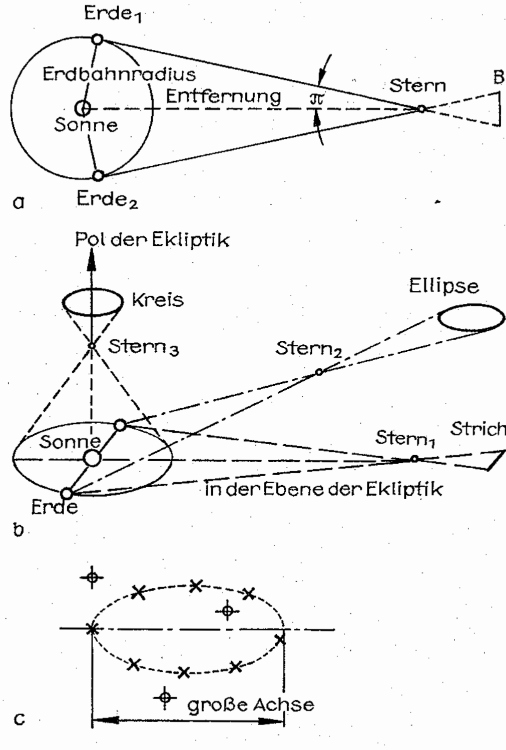


Abb. 7.3.1–8 Jährliche Sternparallaxe; $\sin \pi = a/r \approx \pi$, (a) Blick ⊥ Erdbahn, a = Erdbahnradius, r = Entfernung (π in ″, r in pc), (b) perspektivische Ansicht, (c) Koordinaten von Stern_2 (x) im Laufe eines Jahres gegenüber ortsfesten Bezugssternen, deren $\pi \approx 0$. Aus der großen Halbachse der scheinbaren Ortsveränderung des Sternes folgt die Parallaxe π

messungen von relativ nahen Sternen dient der Erdbahnradius ($AE \approx 150 \cdot 10^6$ km als Basis (Abb. 7.3.1–8). Als Bezugspunkte wählt man mehrere sehr weit entfernte Sterne mit unmeßbar kleiner Parallaxe. Die Messung besteht in etlichen Positionsbestimmungen des betreffenden Sternes im Laufe eines Jahres. Dies geschieht oft mittels fotografischer Aufnahmen bei exakt bekanntem Abbildungsmaßstab. Die aus den Meßpunkten erhaltene Figur besteht allgemein aus einer Ellipse je nach dem Ort des Gestirns bezüglich der Ebene der Erdbahn (die beiden Extremfälle sind ein Kreis oder eine Gerade). Dieses Verfahren ist auf Winkel $\geqq 0{,}01''$ bzw. auf Entfernungen bis 100 pc begrenzt. Kleinere Winkel, d.h. größere Entfernungen, kann man interferometrisch bestimmen (Verschiebung von zwei Spalten vor dem Objektiv eines Refraktors oder zwei räumlich getrennte Instrumente). Auf diese Weise wurden auch Abstände sehr enger Doppelsterne und von einigen genügend nahen Riesensternen die Durchmesser bestimmt. Außer der trigonometrischen Parallaxe mit der Erdbahn als Basis gibt es noch einige andere Verfahren. Dabei dient u. a. ein Teil der Bahn der Sonne um das Zentrum der Galaxis als Basis. Man erhält auch aus den Bewegungen von Sternhaufen wertvolle Hinweise auf ihre Entfernung von der Erde.

7.3.1.6. Verschiedene kosmische Entfernungsbestimmungen. Vergleich der absoluten Helligkeiten der Sterne. Darunter versteht man die Helligkeit eines Sternes, wenn dieser einen Abstand von 10 pc von der Erde hätte. Auf der Erde sieht man die Sterne mit ihrer scheinbaren Helligkeit (je nach ihrer Entfernung, ihrer Größe und Spektraltyp u.a.).

Aus den erwähnten trigonometrischen Sternparallaxen kennt man die relativ genaue Entfernung genügend vieler Sterne der verschiedensten Arten bzw. Spektralklassen, von einigen auch die Durchmesser. Die scheinbaren Helligkeiten kann man fotometrisch bestimmen. Für die Helligkeit gilt die astronomische Größenklassenskale $m_1 - m_2 = -2{,}5 \lg (I_1/I_2)$, m = scheinbare Helligkeit, die von der Erde aus beobachtet wird, I = Intensität. Mit r als Entfernung gilt $I_1/I_2 = (r_2/r_1)^2$ und man erhält $m_1 - m_2 = -5 \lg (r_2/r_1)$. Man setzt nun $r_1 = 10$ pc und entsprechend m_1 = absolute Helligkeit M, so daß $M - m = 5(1 - \lg r)$; r = der gesuchte Abstand in Parsec. m umfaßt Werte zwischen $m = -26$ (Sonne) und $m \approx 24$ (schwächste, mit großen Teleskopen heute erkennbare Objekte).

Die veränderlichen Sterne vom Typ δ-Cephei zeigen einen (fast strengen) Zusammenhang zwischen der Periode der Helligkeitsschwankung und der absoluten Helligkeit. Diese Sterne sind Riesen mit sehr gut bekannten Eigenschaften, vor allem der absoluten Helligkeit. Sie spielen ebenfalls eine Schlüsselrolle bei kosmischen Entfernungsmessungen.

Entfernungen zu anderen Galaxien kann man aus der Rotverschiebung $\Delta\lambda/\lambda$ (λ = Wellenlänge) von Spektrallinien, meist des Kalziums, abschätzen (Doppler-Effekt, vgl. 7.3.4.). Aus $\Delta\lambda/\lambda = [(1 + v/c)(1 - v/c)]^{1/2} - 1$ und $v = 55\,\text{kms}^{-1} \cdot \text{Mpc}^{-1} \cdot r$ (r = Entfernung in Mpc) erhält man die Entfernung (ebenfalls in Megaparsec).

Für kleine Werte $\Delta\lambda/\lambda$ wird $v \approx c \cdot \Delta\lambda/\lambda$. Der Zahlenwert der Konstanten geht natürlich ganz entscheidend in die Betrachtung ein. Sie gibt an, daß (als zur Zeit bester Wert) die Expansionsgeschwindigkeit v der Galaxien je Megaparsec um 55 km/s zunimmt. Auf diese Weise ist die Entfernung r mit der Fluchtgeschwindigkeit v verbunden, die ihrerseits aus der gemessenen Rotverschiebung folgt. Je größer die Entfernung solcher Objekte von der Erde ist, desto höher ist zugleich ihre Fluchtgeschwindigkeit. Die heute höchsten bekannten Werte sind etwa 10^{23} km für die Entfernung und mehr als $c/2$ für die Fluchtgeschwindigkeit.

7.3.1.7. Bestimmung atomarer Dimensionen. Röntgenfeinstrukturanalyse. Bei regelmäßiger Anordnung von Atomen in einer Substanz kann man die Abstände aus der Röntgenfeinstrukturanalyse erhalten.

Aus der Reflexionsbedingung $2a/(h^2 + k^2 + l^2)^{1/2} = \lambda/\sin\Theta$ erhält man die Gitterkonstante a (h, k, l = ganze Zahlen, λ = Wellenlänge der Röntgenstrahlung).

Für hohe Genauigkeitsansprüche werden die Filme fotometrisch ausgewertet und mittels Komparator vermessen. Es müssen möglichst alle Reflexe ausgewertet (wichtig für Gittertyp) und nach richtiger Zuordnung (Indizierung) bei geeigneter Wichtung gemittelt werden. Für viele Gittertypen ist eine Gitterkonstante a sowie auch der Abstand zweier benachbarter Atome von einigen 10^{-10} m charakteristisch. Mit einem leistungsfähigen Elektronenmikroskop kann man solche Abstände (z.B. zwischen zwei Netzebenen eines kubischen Gitters) beobachten und relativ genau bestimmen.

Streuexperimente verschiedener Art (Neutronen, Elektronen, α-Teilchen u.a.) liefern Aufschlüsse über atomare Dimensionen. Erfolgt z. B. nach Abb. 7.3.1–9 die Streuung eines α-Teilchens (Ladung $2e$) an einem schweren Kern (z. B. Gold, sehr dünne Folie) der Ladung Ze, Z = Anzahl der Protonen im Kern, für Gold $Z = 79$, so zeigt die zwischen beiden Teilchen wirkende Coulomb-Kraft (Wechselwirkung zwischen Ladungen) $P \propto 2Ze^2/r^2$ die gleiche Abstandsabhängigkeit (Abstand r) wie die Gravitationskraft $P \propto m_1 m_2/r^2$ (vgl. 7.3.5.). Die Bahn des α-Teilchens ist daher ein Kegelschnitt, in dessen einem Brennpunkt sich der streuende Kern befindet. Da beide Teilchen Ladungen gleichen Vorzeichens tragen, folgt eine Hyperbelbahn nach Abb. 7.3.1–9. Beim zentralen Stoß entartet die Bahn in eine doppelt durchlaufene Gerade mit einem minimalen Abstand d_0 vom Streuzentrum (dort kehrt das α-Teilchen um und fliegt zurück). Es bedeuten weiter: p = geringster Abstand, in dem das Teilchen der Anfangsgeschwindigkeit v_0 ohne Ablenkung am Zentrum vorbeifliegen würde ($p = 0$ heißt dann zentraler Stoß), ϑ = zu p gehörende Ablenkungswinkel (für $p = 0$ wird $\vartheta = 180°$), d_p = der jeweils zu p gehörende kleinste erreichte Abstand des Teilchens vom Kern. d_0 ($p = 0$) erhält man aus der Bedingung, daß dort die ursprüngliche kinetische Energie des α-Teilchens sich in potentielle Energie verwandelt hat. Für ϑ als Funktion des Stoßparameters p ergibt sich $\operatorname{ctg}(\vartheta/2) = 2p/d_0 = p \cdot m \times v_0^2/(2Z\epsilon^2)$, $\epsilon^2 = e^2/4\pi\varepsilon_0$. Die Bahnbeobachtung erfolgt z.B. in einer Nebelkammer und für d_0 erhält man etwa 10^{-14} bis 10^{-15} m, das ist die Ausdehnung des ablenkenden Zen-

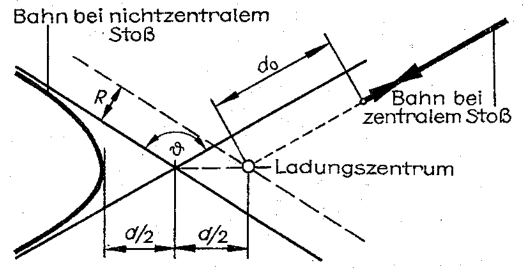


Abb. 7.3.1–9 Streuung eines α-Teilchens an einem Zentrum der Ladung $+Ze$, Bahn bei nichtzentralem und bei zentralem Stoß ($\vartheta = 180°$, $p = 0$)

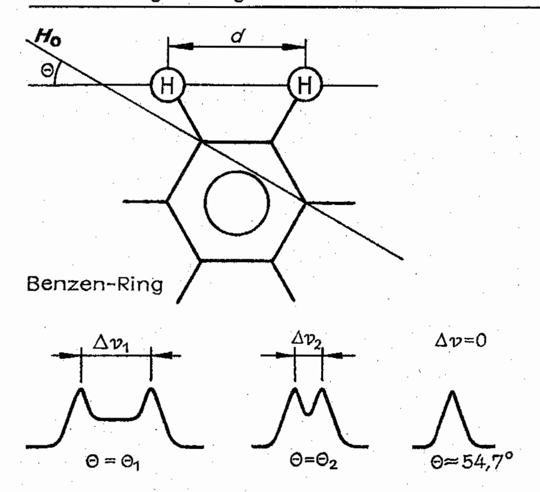


Abb. 7.3.1–10 Aufspaltung einer NMR-Linie durch Dipol-Dipol-Wechselwirkung; H = Wasserstoff (es wechselwirken die beiden Protonen), H_0 Richtung des äußeren Magnetfeldes, $\Delta\nu_{1,2}$ = zu den Winkeln $\Theta_{1,2}$ gehörende Aufspaltungen

trums; dies entspricht etwa der Größe von Atomkernen. In Weiterführung solcher und ähnlicher Experimente erhält man für den Radius eines Atomkernes $r_K = 1{,}3 \cdot A^{1/3} \times 10^{-15}$ m (A = Atomgewicht).

Abstände von Atomen, insbesondere innerhalb von Molekülen gewinnt man in großem Umfang aus spektroskopischen Verfahren, insbesondere der Infrarot (IR)- und der Kernresonanz-Spektroskopie (NMR). Aus der Lage und den Abständen von Rotationslinien (IR) erhält man u. a. das Massenträgheitsmoment I von zweiatomigen Molekülen $I = h/[4\pi^2 c\Delta(1/\lambda)]$ und aus $r^2 = I/\mu$ den Abstand r beider Atome im Molekül ($\mu = m_1 m_2/(m_1 + m_2)$ mit m_1, m_2 = Atommassen). Beispielsweise ergibt sich mit $\Delta(1/\lambda) = 760\ \text{m}^{-1}$ (das ist der Abstand zweier Rotationslinien als Differenz der beiden zugehörigen Wellenzahlen $1/\lambda$) für das CO-Molekül im Grundzustand ein Abstand r beider Atome zu $r = 1{,}13 \cdot 10^{-10}$ m.

Als einfaches Beispiel für die sehr umfangreichen Ergebnisse der Strukturaufklärung mittels der NMR sei die Messung des Abstandes der Wasserstoffatome (2,3) am Benzenring eines Molekülteils unter Berücksichtigung der weiteren Bindungen dieses Ringes (im Molekül) genannt. Infolge der Dipol-Dipol-Wechselwirkung der beiden Protonen (Kerne der H-Atome) kommt es zu einer Aufspaltung $\Delta\nu$ der Resonanzlinie, wobei $\Delta\nu$ vom Winkel Θ zwischen der Verbindungsachse beider Protonen und einem äußeren Magnetfeld $\boldsymbol{H}$ abhängt (Abb. 7.3.1–10). Aus dem (vereinfachten) Zusammenhang $\Delta\nu = \frac{\mu_0}{8\pi^2}\frac{3\gamma^2\hbar}{2r^3}(3\cos^2\Theta - 1)$ und der Grundgleichung der NMR $\nu = (\gamma/2\pi)B$. (γ = gyromagnetisches Verhältnis des betreffenden Atomkerns) erhält man für das spezielle Beispiel einen Abstand r der beiden H-Atome (Mitte/Mitte) zu $r \approx 2{,}4 \cdot 10^{-10}$ m. Dieses Verfahren ist sehr empfindlich und gestattet den Nachweis von sehr kleinen Abstandsänderungen und gibt damit wertvolle Hinweise u. a. auch auf Änderungen von Bindungswinkeln. Durch gezielte Substituierung von Wasserstoff durch Deuterium lassen sich in weiten Grenzen beliebige Molekülteile untersuchen (Deuterium anstelle von Wasserstoff zeigt gleiches chemisches Verhalten, aber völlig anderes γ).

7.3.1.8. Meßwandler zur Messung von Längenänderungen. Für eine Reihe von Meßverfahren benötigt man Meßwandler zur Umsetzung von mechanischen Größen in elektrische Größen. Diese Meßwertwandler haben auch dort Bedeutung, wo z. B. Kräfte (Drücke) über Längenänderungen nachgewiesen werden.

Aus der großen Zahl von Möglichkeiten sollen einige elektrische Wandler für Längenänderungen erwähnt werden.

Dehnungsmeßstreifen zeigen die Widerstandsänderung des Streifenmaterials bei Längenänderung an. Der Streifen ist mit dem Testkörper sicher verklebt. Die Widerstände R haben meist Werte zwischen 500 und 1 000 Ω. Die Anordnungen werden i. allg. mit Wechselspannungen (500 bis 1 000 Hz) betrieben, die meßbaren Werte für $\Delta R/R$ sind $\leq 10^{-4}$; es lassen sich Längenänderung Δl bis in den Mikrometerbereich erfassen.

Überträgt man die Längenänderung Δl auf die bewegliche Elektrode eines Kondensators, so dient die *Kapazitätsänderung* ΔC zum Nachweis. Dazu verwendet man entweder Kapazitätsmeßbrücken oder ordnet den Kondensator in einem Schwingkreis an und mißt dessen Änderung $\Delta\nu$ der Resonanzfrequenz $\nu = 1/2\pi \cdot (L \cdot C)^{-1/2}$; $|\Delta\nu| \approx 1/4\pi \cdot L^{-1/2} \times C^{-3/2}\Delta C$. Es lassen sich je nach Frequenz und Anordnung Werte von $\Delta l \leq 0{,}1$ µm nachweisen. Bei Übertragung dieses Verfahrens auf Hohlraumresonatoren (Frequenzen im GHz-Bereich) sind bei extremen Anforderungen an Meßwandler, Umgebungsbedingungen und Elektronik Werte Δl bis zu 10^{-8} m und weniger erreichbar.

Wird für die Messung von Längenänderungen ein *induktiver Meßwandler* verwendet, kann eine solche Anordnung auch als Kraftmeßdose dienen, solange die durch Kräfte verursachte Deformation der Membran innerhalb der Proportionalitätsgrenze verbleibt (vgl. 7.3.5.).

7.3.1.9. Flächen und Volumina. Flächen und Volumina werden bei geometrisch behandelbaren Formen auf Längen- und Winkelmessungen zurückgeführt. Ist dies nicht möglich, kann man Flächen unter Umständen mit einem Planimeter bestimmen, oder man teilt die betreffende Fläche in Quadrate (Dreiecke o. ä.) und zählt diese unter Beachtung einer erforderlichen Mittelung aus (grafische Integration). Man kann auch die betreffende Figur auf dünnen Karton zeichnen, ausschneiden und wägen. Ein genau vermeßbares Stück des gleichen Materials wird ebenfalls gewogen und die gesuchte Fläche ergibt sich als Verhältnis beider Wägungen. Das Volumen eines unregelmäßigen Körpers der Masse m und bekannter Dichte ϱ folgt aus einer Wägung ($V = m/\varrho$, vgl. 7.3.2.). In anderen Fällen taucht man den Körper unbekannten Volumens in eine (mit der Substanz verträgliche) Flüssigkeit und mißt die dabei verdrängte Flüssigkeitsmenge (Meßzylinder, Überlaufmethode o. ä.), die gleich dem des gesuchten Volumens ist.

7.3.2. Bestimmung von Massen und Dichten

7.3.2.1. Bestimmung der Masse. Waagen dienen in erster Linie zur Bestimmung von Massen. Als Waage bezeichnet man jedes Instrument, das die von einer Masse auf eine Unterlage ausgeübte Kraft durch Vergleich mit einer gleichen Masse oder mit einer beliebigen Meßanordnung für diese Kraft zu messen gestattet. *Hebelwaagen* erlauben die fast kräftefreie Umlenkung der Kompensationskraft. Die

„klassische" Labor-Hebelwaage mit zwei Schalen wurde zu hoher Vollkommenheit entwickelt. Für Massen bis zu etwa 200 g beträgt die Genauigkeit wenigstens 0,1 mg. Die Empfindlichkeit dieser Waagen ist von der Belastung abhängig. Da bei genauen Wägungen außerdem das Gleichgewicht aus den aufeinanderfolgenden Nulldurchgängen des Zeigers (am Waagebalken) bis zum Stillstand bestimmt werden muß, sind solche Messungen oft zeitaufwendig. Die verwendeten Gewichtssätze bedürfen einer sehr sorgfältigen Behandlung und öfteren Prüfung „in sich". Man strebt für Laborwaagen maximale Empfindlichkeit und minimale Meßzeit an. Diese Überlegungen führten zur *Einschalenwaage*, bei der die zu wägende Masse und die Gewichte auf einer Seite des Waagebalkens wirken, während der andere Waagebalken ohne Last für den gesamten aufgelegten Gewichtssatz das Gleichgewicht hält. Bei Auflegen einer Masse auf die Schale wird durch einen Gewichtsauflegemechanismus soviel abgehoben, bis die Waage im Gleichgewicht ist. Dieser Mechanismus wird über Schaltknöpfe, z.B. 10 bis 190 g, 1 bis 9 g und 10 bis 990 mg betätigt und eine Projektionsskala für den Bereich 0 bis 10 mg gestattet eine rasche Ablesung auf etwa 0,2 mg für eine Routinewaage. Die beiden Waagebalken sind hier immer mit der gleichen maximalen Last belegt, womit die Durchbiegung der Waagebalken und damit die Empfindlichkeit konstant bleibt. Die besten Waagen dieser Art messen auf 1/500 mg genau. Der Waagenmechanismus kann auch elektrisch oder magnetisch betätigt werden. Hierbei erfolgt die Wägung selbstkompensierend (automatisch). Für sehr genaue Bestimmung kleiner Massen (0,1...max 1 g) werden hochempfindliche *Mikrowaagen* eingesetzt. Die genauesten bekannten Wägungen wurden auf 10^{-9} bis 10^{-10} genau ausgeführt (Schwingungszeit bis zu einigen Minuten, Balkenmasse 0,4 g, Last maximal 0,2 g). Bei genauen Wägungen ist der unterschiedliche Auftrieb der zu bestimmenden Masse und der benutzten Gewichtsstücke zu korrigieren. Es gilt $M = G + \varrho_L(v_m - v_G) = G(1 + \varrho_L/\varrho_m - \varrho_L/\varrho_g)$, M = zu wägende Masse, G = Gewichtsstück(e) der Masse G, v_m = Volumen der Masse M, ϱ_m = Dichte der Masse M, v_G = Volumen der Gewichtsstücke, ϱ_G = Dichte der Stücke G, ϱ_L = Luftdichte. Diese ist eine Funktion der Temperatur T, des Druckes p und des Wasserdampfanteils. Im Mittel gilt 1,293 kg/m³ Luft (genau auf 10^{-3} bis 50 % Luftfeuchtigkeit).

Bei einer Zweischalenwaage sind manchmal Wägungen mit vertauschten Schalen erforderlich (Mittelbildung). Messungen im Vakuum sind technologisch schwierig zu beherrschen und zeigen oft erhebliche Schwierigkeiten (bestimmte Oberflächenschichten beim Evakuieren u. a.).

Federwaagen beruhen auf der Wirkung der Kraft P einer Masse M auf die Dehnung der Feder innerhalb der Proportionalität zwischen Kraft und Dehnung (Abb. 7.3.2–1a). Bei einer Dehnung von 40 m/g ist mit dem Mikroskop eine Ablesung auf weniger als 1 mg möglich. Federwaagen können je nach Konstruktion für einen sehr großen Kräftebereich benutzt werden.

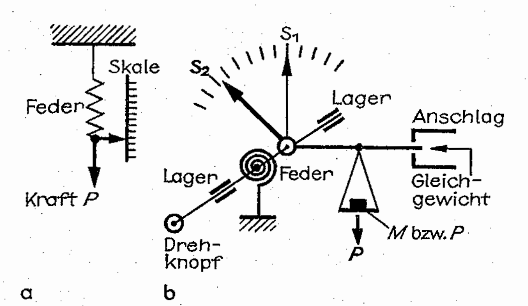


Abb. 7.3.2–1 (a) Federwaage, (b) Torsionswaage

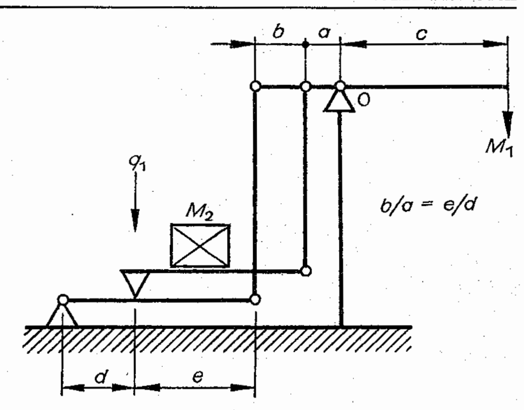


Abb. 7.3.2–2 Prinzip einer Dezimalwaage; Gleichgewicht: $M_1 c = M_2 a$ (Parallelverschiebung der Last-Platte c/a = 1:10 [Dezimal] oder 1:100 o.ä.)

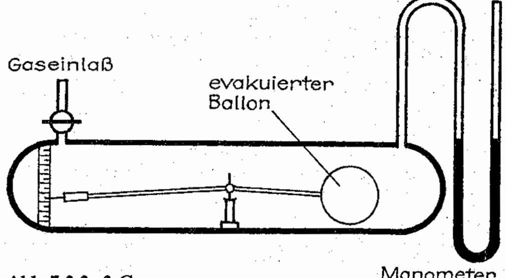


Abb. 7.3.2–3 Gaswaage

Abb. 7.3.2–1b zeigt das Prinzip einer *Torsionswaage*. Mittels der Feder wird die Waage nach Auflage der Last wieder in die Stellung „Null" gebracht. Die dazu erforderliche Verdrehung der Feder liest man an der Skale ab. Als unterste Meßgrenze wird 2 mm Skalenlänge $\hat{=} 10^{-6}$ g angenommen.

Für größere Massen verwendet man die *Dezimalwaage* (Hebelverhältnis 1:10 und 1:100, Abb. 7.3.2–2).

Die Abb. 7.3.2–3 zeigt eine *Gaswaage*. Der Ballon ist gut evakuiert und erhält je nach der Gasdichte des zu messenden Gases einen unterschiedlichen Auftrieb, der an der Skale angezeigt wird. Der ganze Kolben mit der Waage wird jeweils evakuiert, mit dem Meßgas gespült und dann bis zum Druck p gefüllt ($p < \varrho \cdot g_E$, ϱ = Gasdichte).

Zur Bestimmung der Massen von Molekülen oder Atomen (oder Elementarteilchen) wendet man indirekte Meßverfahren an, die aber stets an direkte Verfahren, d.h. an die definierte Masseneinheit anschließen. Einige solcher Verfahren sollen im folgenden als Beispiele dienen.

In einem der einfachsten Fälle kann man bei geeigneten Bedingungen die Molekül- bzw. Atommasse aus der Dampfdichte (vgl. 7.3.2.2.) erhalten. Aus der Beugung von Röntgen-Strahlen kann man in geeigneten Fällen die Gitterkonstante und die Art des Gitters entnehmen. Bei Silber befinden sich (als Ergebnis dieser Messung) in einem Wür-

fel der Kantenlänge $4{,}08 \cdot 10^{-10}$ m vier Ag-Atome. Die makroskopisch ermittelte Dichte ϱ betrage $10{,}5 \cdot 10^3$ kg/m³. Für die Masse eines Ag-Atoms folgt
$m_{Ag} = 10{,}5 \cdot 10^3\,\mathrm{kg/m^3} \cdot (4{,}08 \cdot 10^{-10})^3\,\mathrm{m^3} \cdot 1/4 = 1{,}79 \cdot 10^{-25}$ kg. Damit ergibt sich auch das (chemische) Atomgewicht (als Mittel der beiden stabilen Ag-Isotope) zu etwa 108.

Massenspektroskopie ist eine allgemeine Methode der Bestimmung der Massen von Atomen, Molekülen oder charakteristischen Teilen davon sowie der Isotopenzusammensetzung. Einem Massenspektrometer wird die zu untersuchende Substanz gas- bzw. dampfförmig zugeführt oder im Gerät verdampft. Die Teilchen werden u. a. durch Elektronenstoß ionisiert und können so mittels elektrischer und/oder magnetischer Felder oder Laufzeiteffekten entsprechend dem Verhältnis Ladung/Masse (e/m) getrennt und separat vom Detektor aufgefangen und registriert werden. Diese Linien der betreffenden Massenzahlen werden auf die Masse des Kohlenstoffes ^{12}C mit der Massenzahl 12,000 00 bezogen. Mit der atomaren Masseneinheit $u = 1{,}66\; 10^{-27}$ kg erhält man die absolute Atom- oder Molekülmasse bzw. deren Bruchteile. Die Abb. 7.3.2–4 zeigt zwei Ausschnitte des Massenspektrums von $CHCl_3$ (bei niedriger Auflösung). Man erkennt die Linien der Massenzahlen 35 und 37, die unter Berücksichtigung der übrigen Massenzahlen entsprechend der natürlichen Häufigkeit von 75,5:24,5 eindeutig den beiden Chlorisotopen ^{35}Cl und ^{37}Cl zugeordnet werden können. Die Massenzahlen 47 und 49 gehören zu $C^{35}Cl$ und $C^{37}Cl$. Der massenspektrometrische Molekülpeak wird entsprechend $^1H^{12}C^{35}Cl_3$ zu M = 118 identifiziert. Der Peak 120 gehört zu $HC^{35}Cl_2{}^{37}Cl$ und M = 122 zu $HC^{35}Cl^{37}Cl_2$. Die anderen Massenlinien gehören zu Bruchstücken, die bei der Ionisierung entstehen und für die Massenspektroskopie charakteristisch sind. Entsprechend der geringen Auflösung des verwendeten Gerätes erhielt man als Masse des Isotops ^{35}Cl demnach $35\ u = 58{,}1 \times 10^{-27}$ kg (der genaue Wert beträgt $58{,}05797 \cdot 10^{-27}$ kg). Das chemische Atomgewicht als Mittel beider Isotope entsprechend ihrer Häufigkeiten 0,755:0,245 beträgt rund $35 \times 0{,}755 + 37 \cdot 0{,}245 = 35{,}5$.

Die Masse des Elektrons kann man u. a. auch aus dem Verhältnis Ladung/Masse (e/m) bestimmen. Passiert ein Elektron mit der Geschwindigkeit $v \ll c$ ein elektrisches Feld $\boldsymbol{E}$, so bewirkt die Kraft $\boldsymbol{P} = e\boldsymbol{E}$ eine Auslenkung gegen die ursprüngliche (geradlinige) Flugbahn. Diese Ablenkung ist auf einem nachgestellten Schirm meßbar. Da v, E und e bekannt sind, kann man m ermitteln.

Massen von Himmelskörpern bestimmt man in erster Linie aus dem Gravitationsgesetz $P = \gamma \cdot m_1 m_2 / r_{1,2}^2$ und dem 3. Kepler'schen Gesetz $a^3/T^2 = \text{const.}$ (r = Abstand beider Massenmittelpunkte, a = große Achse, T = Umlaufzeit). Die Graviationskonstante γ kann nicht den Bewegungsgesetzen der Himmelsmechanik entnommen werden, sondern mußte in sehr sorgfältigen Messungen im irdischen Labor ermittelt werden (man muß die Kraft zwischen zwei bekannten Massen bestimmen). Als Ergebnis erhielt man $\gamma = 6{,}67 \cdot 10^{-11}\,\mathrm{N \cdot m^2/kg^2} = 6{,}67 \cdot 10^{-11}\,\mathrm{kg^{-1} \cdot m^3 \cdot s^{-2}}$. Mit der Kenntnis von γ und der Erdbeschleunigung $g_E = 9{,}80665\,\mathrm{m/s^2}$ (Normwert) folgt aus $mg_E = \gamma \cdot m \cdot m_E/r^2$ die Erdmasse $m_E \approx 9{,}8\,\mathrm{m \cdot s^{-2}} (6{,}37 \cdot 10^6)^2\,\mathrm{m^2}/6{,}67 \times 10^{-11}\,\mathrm{kg^{-1}\,m^3\,s^{-2}} = 5{,}97 \cdot 10^{24}$ kg. Die Masse der Erde wurde also aus einer Kraftmessung (γ), einer Zeit- und Längenmessung (g) sowie Winkel- und Längenmessung (r = Erdradius) bestimmt.

Die Masse der Sonne folgt aus dem Gleichsetzen der Zentrifugalkraft der Erde auf ihrer Bahn um die Sonne und der

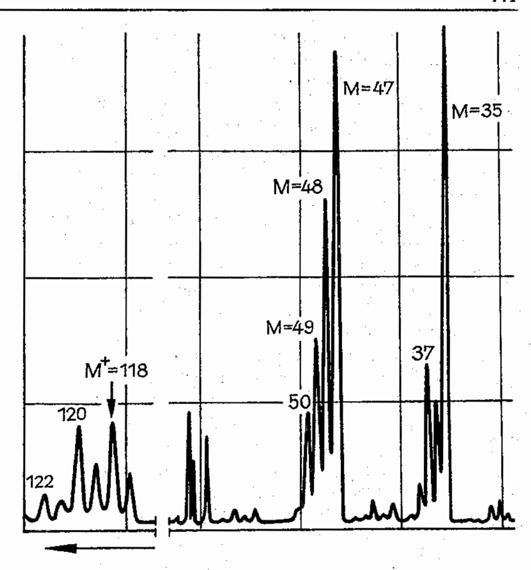


Abb. 7.3.2–4 Teile des Massenspektrums von $CHCl_3$, massenspektrometrischer Molekülpeak $M^+ = 118$

Gravitation zwischen Sonne und Erde zu $\gamma m_E m_s / r^2 = m_E \times \omega^2 \cdot r = m_E \cdot v^2/r$, dabei sind ω = Winkelgeschwindigkeit der Erdbahn $\approx 2\pi/31{,}5 \cdot 10^6\,\mathrm{s^{-1}}$, v = Bahngeschwindigkeit der Erde um die Sonne $\approx \omega \cdot r$, r = Abstand Sonne–Erde = Erdbahnradius $= 1{,}495 \cdot 10^{11}$ m, $m_{Sonne} = v^2 \cdot r/\gamma \approx 2 \times 10^{30}$ kg.

Massen von Fixsternen werden nach mehreren Methoden abgeschätzt. Zwischen Spektralklasse, absoluter Helligkeit, Masse, Radius, effektiver Temperatur u. a. bestehen in vielen Fällen eindeutige Beziehungen, aus denen die in das Spektralklassen-Leuchtkraft-Diagramm eingezeichneten Angaben über Masse, Temperatur und Radius (in jeweiligen Einheiten der Sonne) folgen.

Eine relativ einfache Methode zur Bestimmung der Massen von Doppelsternen besteht in der Beobachtung des Lichtwechsels (Periode T) beim gegenseitigen Umlauf (Bedekkungsveränderliche) beider Komponenten ($T \mathrel{\hat=}$ Zeitmessung) sowie in der Bestimmung ihres Abstandes (große Achse $a \mathrel{\hat=}$ Winkelmessung). Die dazu notwendige Entfernung des Doppelsterns von der Erde erhält man in vielen derartigen Fällen aus anderen, aber gut abgesicherten Überlegungen. Setzt man diese Werte in das 3. Kepler'sche Gesetz in seiner vollständigen Form ein, so folgt aus $a^3/T^2\,(m_1 + m_2) = \gamma/4\pi^2$ die Gesamtmasse $(m_1 + m_2)$ des Systems. Es ist oft möglich, auch auf das Verhältnis m_1/m_2 zu schließen.

Eine Abschätzung der *Masse des Milchstraßensystems* folgt aus der Rotation um sein Zentrum, an der auch die Sonne teilnimmt (Abstand Zentrum–Sonne $\approx 3 \cdot 10^{20}$ m, Bahngeschwindigkeit $\approx 2{,}5 \cdot 10^5$ m/s, Umlaufdauer $\approx 8 \cdot 10^{15}$ s; Werte aus Winkelmessungen). Unter Berücksichtigung der Dichteverteilung liegt der weitaus größte Teil, insbesondere das gesamte massereiche Zentrum der Galaxis, innerhalb der Sonnenbahn. Man erhält daraus für die Milchstraße eine Masse von $\approx 4 \cdot 10^{41}\,\mathrm{kg} = 2 \cdot 10^{11}$ Sonnenmassen.

7.3.2.2. Dichtebestimmung. Die Dichte ϱ ist die in der Volumeneinheit enthaltene Masse, d. h. $\varrho = m/V$. Die direkte Messung geht auf Massenbestimmung (oft Wägung) und

Volumenbestimmung zurück. Die Dichte fester Körper erhält man u. a. durch Wägemethoden. Bei der Auftriebsmethode (hydrostatische Waage, Senkwaage) wird eine Wägung in Luft und in einer Flüssigkeit bekannter Dichte ϱ_{Fl}, meist Wasser, kombiniert. In Luft gewogen und auf Vakuum korrigiert betrage das Gewicht p' und unter Wasser p, der Auftrieb ist dann $(p' - p)$ und man erhält die Dichte zu $\varrho = (p'/(p' - p))\varrho_{Fl}$. Die Dichte ϱ von Substanzen, die in nur kleinen Stücken oder Pulver vorliegen, läßt sich unter Umständen messen, indem man diese in einer Mischung von zwei Flüssigkeiten der Dichten ϱ_{F1} und ϱ_{F2} zum Schweben bringt (wegen der Oberflächenspannung deutlich unter der Oberfläche). Dazu mischt man z. B. Methyleniodid ($\varrho_{F1} \approx 3{,}3$ und Xylol $\varrho_{F2} \approx 0{,}88$), bis die Dichte der Mischung $\varrho_{Mi} = \varrho$. Diese Methode ist begrenzt auf $\varrho \leqq 3{,}3 \times 10^3$ kg/m³. Anschließend bestimmt man die Dichte der Mischung. Die Wägemethoden umfassen u. a. solche mit Meßkolben, Pipetten, Büretten sowie *Pyknometern*. Letztere sind Wägefläschchen bekannten Inhalts mit einer Marke oder Skale am Gefäßhals. Eine zweckmäßige Ausführung mit Thermometer und zwei Öffnungen (eine zum Einfüllen, die zweite zum eventuell notwendigen Absaugen der Luft) zeigt Abb. 7.3.2–5. Die Volumenänderung mit der Temperaturänderung ist zu beachten. Es sind Genauigkeiten unter 0,1 ‰ erreichbar. Zur schnellen angenäherten Bestimmung von ϱ (von Flüssigkeiten) verwendet man *Aräometer*. Sie bestehen aus einem Schwebekörper mit Skale, wobei die jeweilige Eintauchtiefe in die zu messende Flüssigkeit ein Maß für deren Dichte ist. Die Kalibrierung erfolgt in Flüssigkeiten genau bekannter Dichte. Bei den Wägemethoden sind die Ergebnisse, wenn nötig, auf das Vakuum und Wasser von 4 °C zu reduzieren.

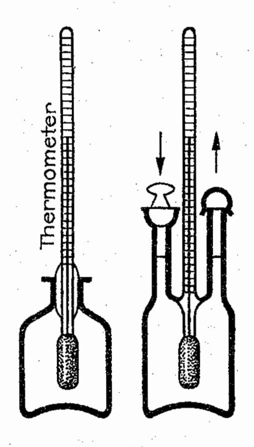


Abb. 7.3.2–5 Pyknometer

Die *Gasdichte* bzw. *Dampfdichte* bestimmt man u. a. mit der Gaswaage (vgl. Abb. 7.3.2–3). Wegen der starken Abhängigkeit der Gasdichte von der Temperatur und dem Druck werden diese Dichten im Normalzustand ($p_0 = 1{,}013 \times 10^5$ Pa, $T_0 = 273{,}16$ K $\triangleq$ 0 °C) angegeben. Wählt man Luft als Bezugsgas, muß man für genaue Messungen auch den Wasserdampfgehalt sowie den Gehalt an Kohlensäure berücksichtigen.

Die Dichten verschiedener Gase verhalten sich nahezu umgekehrt wie die Quadrate der Ausströmgeschwindigkeiten, mit denen die Gase unter der Einwirkung gleicher Drücke aus engen Öffnungen austreten. Die Ausströmzeit t einer bestimmten Gasmenge ist umgekehrt proportional der Ausströmgeschwindigkeit, daher gilt $\varrho_1/\varrho_2 = t_1^2/t_2^2$ (ϱ_1 oder ϱ_2 muß bekannt sein).

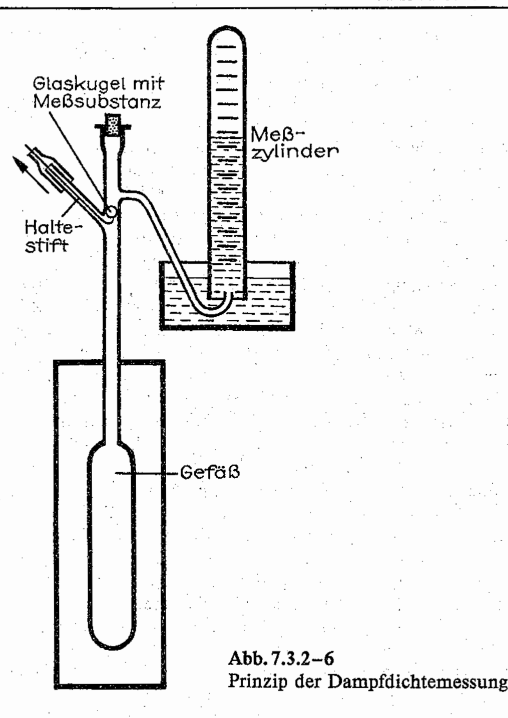


Abb. 7.3.2–6
Prinzip der Dampfdichtemessung

Eine Möglichkeit zur Dampfdichtemessung ist in Abb. 7.3.2–6 skizziert. Die Apparatur ist mit Luft oder einem anderen Gas unter Normaldruck gefüllt. Wenn sich im Kolben eine konstante Temperatur eingestellt hat, läßt man durch Herausziehen des Haltestiftes eine dünnwandige Glaskugel, in der sich eine Probe der zu untersuchenden Substanz (flüssig) befindet, in das Gefäß fallen. Dabei zerplatzt die Kugel, und die Substanz verdampft rasch unter der Voraussetzung, daß die Temperatur im Gefäß erheblich über der Verdampfungstemperatur der Substanz liegt. Die verdampfende Substanzmenge m verdrängt ein gleiches Volumen Luft, die mittels eines Meßzylinders aufgefangen und bestimmt wird. Man erhält die Dampfdichte D auf einige Prozent genau zu $D = (m/V)\ (P_0/P)\ (1 + 0{,}004\,t)$, wobei m = Substanzmenge (kg), V = im Meßzylinder aufgefangenes Volumen (m³), p_0 = Normaldruck $= 1{,}013 \cdot 10^5$ Pa, p = Luftdruck (Pa) − $9{,}8 \cdot h$ (mm), t = Temperatur (°C). Der Faktor 0,004 berücksichtigt die Sättigung der einzelnen Apparateteile (für ideale Gase ist der Ausdehnungskoeffizient 0,003 7/Grad).

Dichte im Atomkern. Sie ergibt sich z. B. aus dem Protonenradius ($\approx 0{,}8 \cdot 10^{-15}$ m) und seiner Masse ($\approx 1{,}66 \cdot 10^{-27}$ kg) zu etwa $0{,}75 \cdot 10^{18}$ kg/m³.

Mittlere Dichte $\bar{\varrho}$ von Himmelskörpern. Sie folgt aus Masse und Volumen (über den Radius). $\bar{\varrho}$ der Sonne ergibt sich aus direkten Messungen zu $\approx 1{,}46 \cdot 10^3$ kg/m³. Einige Beziehungen zwischen den Zustandsgrößen von Sternen wurden bereits erwähnt (vgl. 7.3.2–1). Die Leuchtkraft L ist mit dem Radius R über das Strahlungsgesetz $L = 4\pi R^2 \sigma T_e^4$, ($\sigma = 5{,}7 \cdot 10^{-8}$ W · m⁻²K⁻⁴) verknüpft. Die Oberflächentemperatur T_e folgt in erster Linie aus der Intensitätsverteilung der Wellenlängen des emittierten Lichtes. Ein Beispiel für

hohe Dichten in Sternen erhält man aus den (kleinen) periodischen Bewegungen des Sirius A (Hauptreihenstern von etwa zweifacher Sonnenmasse, $T_e = 9\,250$ K; Entfernung und Radius aus Parallaxe bestimmbar), die auf einen Begleiter (Sirius B) von knapp einer Sonnenmasse schließen lassen. Dieser hat aber (mit $T = 8\,200$ K) nur 1/10000 der Helligkeit des Sirius A, muß also sehr klein sein. Sein Radius ergab $R = 1{,}8 \cdot 10^7$ m und somit $\bar{\varrho} \approx 7{,}3 \cdot 10^7$ kg/m³ (entspricht weißen Zwergen). Das Ergebnis wird auch durch eine (geringe) Rotverschiebung der H_α-Linie von etwa $\Delta\lambda/\lambda \approx 0{,}8 \cdot 10^{-4}$ gestützt, die nach $\Delta\lambda/\lambda \approx \Delta\varphi/c^2$ auf ein hohes Graviationspotential ($\varphi = -\gamma m/R$), d. h. auch kleines R, hinweist ($\Delta\varphi = \varphi_{Stern} - \varphi_{Erde}$). Man muß hier sicher sein, daß zu $\Delta\lambda/\lambda$ keine Anteile aus Bewegungsvorgängen beitragen, und das ist eine schwer zu erfüllende Forderung.

7.3.3. Zeit- und Frequenzmessung

7.3.3.1. Zeitmessung. Hierzu bedarf es der Festlegung eines reproduzierbaren und unveränderlichen Zeitmaßes. Physikalische Zeitmessung beruht auf Konvention, die als Definition eingeführt wird. Dazu sind prinzipiell streng periodische Vorgänge geeignet (Zeitmaß ≙ Periodendauer). Seit 1967 gilt als das stabilste Zeitmaß, die SI-Sekunde (s), definiert als die Dauer von 9192631770 Perioden der Strahlung, die dem Übergang zwischen den beiden Hyperfeinstrukturniveaus $F = 4$, $m_F = 0$ und $F = 3$, $m_F = 0$ des Grundzustandes $^2S_{1/2}$ des Atomes Caesium 133 (^{133}Cs) entspricht. Das Funktionsprinzip dieser Atomuhren beruht auf der Wechselwirkung der genannten Atomübergänge mit magnetischen Hochfrequenzfeldern, die von einem Hilfsgenerator gespeist werden. Bei Übereinstimmung beider Frequenzen zeigt die am Detektor registrierte Intensität des Atomstrahles ein Minimum, auf das mittels Abtastverfahren (Frequenzmodulation des Trägers) eine exakte (elektronische) Stabilisierung (z.B. phasenempfindliche Regelspannung) des Hilfsgenerators möglich ist. Die Grenze der dabei erzielbaren Genauigkeit wird durch die Schärfe des Minimums und die zur Funktion erforderliche minimale Regelabweichung gebildet. Bei den besten Konstruktionen liegt die Breite des Regelminimums bei $\approx 10^{-3}$ Hz (und weniger), was bei $\nu \approx 10^{10}$ Hz einer Unsicherheit von etwa 10^{-13} (als Untergrenze gilt heute 10^{-14}) entspricht. In weniger günstigen Fällen beträgt $\Delta\nu/\nu \approx 10^{-11}$. Zwei Atomuhren (gleicher Güte) zeigen also im besten Fall mit den genannten Werten innerhalb eines Jahres ($\approx 31{,}5 \cdot 10^6$ s) einige 10^{-6} gegenseitige Abweichung, d. h. eine solche Uhr ginge erst in etwa einigen 10^5 bis 10^6 Jahren um eine Sekunde falsch. Der für die Abtastung der Cs-Resonanz erforderliche Hilfsgenerator kann z. B. ein Quarzoszillator sein. Durch Frequenzvervielfachung und Mischung wird eine dem Cs-Übergang entsprechende Frequenz synthetisiert, die dann, wie beschrieben, auf die Cs-Resonanz stabilisiert wird. Durch Teilung der Frequenz des Oszillators werden verschiedene Normalfrequenzen und Sekundenmarken gewonnen, deren Folge die Atomzeitskale mit einem Höchstmaß an Gangstabilität darstellt (*primäres Normal*).

7.3.3.2. Sekundäre Zeitnormale. Sie werden als Uhren für die Interpolation und Extrapolation aufgrund von Zeitpunkten, die durch das primäre Normal festgelegt sind, benutzt. Uhren aller Art enthalten einen Taktgeber, eine Zählvorrichtung für die Taktperioden und eine Anzeigevorrichtung. Zum Vergleich zweier Taktgeber (fast gleicher Taktfrequenzen), z. B. periodischer Impulse, mißt man zweckmäßig deren gegenseitige Phasenbeziehung, wobei das Normal den Bezugswert liefert. In einem einfachen Fall gibt man z. B. beide Impulsfolgen auf einen Oszillografen und vergleicht bzw. mißt die entstehenden zeitlichen Abstände der Impulse (Abb. 7.3.3–1). Dieses Verfahren wird auch zur Regulierung mechanischer Uhren angewendet, wobei mittels geeigneter Meßwandler der mechanische Takt in elektrische Impulse umgeformt wird. Zeigt der Prüfling z.B. nach 100 Takten (s) eine Abweichung von 10^{-3} s (leicht meßbar), so ist der Fehler je Sekunde demnach 10^{-5} (diese Uhr ginge innerhalb eines Tages ≈ 1 s falsch). Mittels moderner Impulsmeßtechnik kann man einmalige Vorgänge auf 10^{-9} s/mm und periodische Vorgänge auf $\approx 10^{-11}$ s/mm Oszillogrammlänge ausmessen. Auch mit direkten Phasenmessungen kann man Frequenzvergleiche sehr hoher Genauigkeit durchführen.

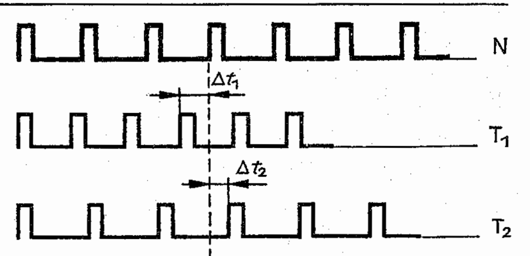


Abb. 7.3.3–1 Phasenvergleich von Impulsfolgen (Taktfrequenzen), T_1: Taktgeber hat höhere Frequenz als Normal N; T_2: Taktgeber hat niedrigere Frequenz als Normal N

7.3.3.3. Internationale Atomzeit (IAT). Sie wird nach internationaler Übereinkunft aus einer Anzahl von Atomuhren mittels geeigneter Mittelung festgelegt (SI-Sekundendefinition). Die extrem hohe Genauigkeit dient wissenschaftlichen Zwecken und bildet die Grundlage für Normalfrequenzen auf 10^{-10} genau. Aus astronomischen Ereignissen folgt eine *Weltzeit UT* (*u*niversal *t*ime), die aus der Erdrotation abgeleitet wird und die man nach Korrekturen infolge von Polbewegungen der Erde als UT 1 bezeichnet. Für astronomische Beobachtungen sowie Navigation nach Himmelskörpern ist UT 1 maßgebend, die aber infolge verschiedener Einflüsse mit einer Unsicherheit von etwa 10^{-7} behaftet ist. Aus dem Vergleich der Atomzeit (IAT) und UT 1 wird nach (internationaler) Vereinbarung die *koordinierte Weltzeit UTC* definiert, die Eigenschaften der Atomzeit hat (nämlich bezüglich der Gleichmäßigkeit des Sekundentaktes), aber langfristig gegen UT 1 eine Abweichung (DUT 1) bis max. 0,9 s haben darf. Die Differenz UTC − UT 1 = DUT 1 wird nachträglich in 0,1-s-Schritten bekannt gegeben (im nationalen Rahmen können bei Bedarf noch Unterteilungen bis 0,02 s als DUT 1) erfolgen. Erreicht DUT 1 eine Sekunde, so wird eine Schaltsekunde (allgemein etwa im Abstand eines halben Jahres) eingeschoben. Die letzte Minute des letzten betreffenden Monatstages hat dann 61 s (oder auch nur 59 s). Für die bürgerliche Zeitrechnung, die auf der Zeitskale UTC beruht, benutzt man also, da nicht höchstmögliche Genauigkeit über lange Zeiträume verlangt wird, die mittlere Sonnenzeit, der der mittlere Sonnentag zu 24 Stunden = 86400 Sekunden zugrunde liegt. Der Beginn der Tageszählung fällt für einen Ort bestimmter geografischer Länge auf die untere Kulmination der mittleren Sonne im Meridian des Beobachtungsortes. 00.00 h UTC bezieht sich auf die geografische Länge 0 Grad (Greenwich). Die Zeitzonen der Erde sind auf UTC bezogen, z.B.

ist die mitteleuropäische Zeit (MEZ) = UTC + 1 h. UTC wird mittels *Zeitsignale* mit einer Genauigkeit von 10^{-4} s über bestimmte Sender mit bekannten Koordinaten zu vereinbarten Zeiten abgestrahlt. Beim Empfang sind die Laufzeiteffekte zu beachten (je 300 km Entfernung 10^{-3} s). Bei der Positionsbestimmung von Himmelskörpern beträgt deren scheinbare Bewegung infolge der Erdrotation 15″/s. Eine (gut meßbare) Winkeldifferenz von 0,1″ wird demnach in $6{,}7 \cdot 10^{-2}$ s durchlaufen, so daß i. allg. ein absoluter Zeitvergleich auf 10^{-4} s für den genannten Zweck als ausreichend erscheint.

Die Angabe von UTC kann auch mittels der Synchronimpulse des Fernsehens durch nachträgliche Bekanntgabe des zeitlichen Abstandes des ersten Impulses von einem vorher vereinbarten Zeitpunkt UTC für jeden Tag erfolgen. Beispiel: Für den betreffenden Impuls vereinbarter Zeitpunkt UTC = $13^h\,15^{min}\,00^s$; zum Zeitpunkt der Aussendung festgestellte (aber nachträglich bekanntgegebene) Differenz = +0,018 119 6 s; die Impulsflanke entsprach daher am Ort des Senders (der Antenne) einer Angabe UTC = $13^h\,15^{min}\,0{,}018\,119\,6^s$. Die auf den Sendeort reduzierte Zeitskale am Ort des Empfängers kann damit nachträglich auf 10^{-7} korrigiert werden.

Die beim Fernsehen verwendete Zeilenfrequenz 15 625 Hz (mal $2^6 = 10^6$) dient auch für Meßzwecke als Normalfrequenz. Es wird ein Zeitintervall (z. B. täglich 10 min. zu einem bestimmten Zeitpunkt) vereinbart, innerhalb dessen die Abweichungen von der durch die Atomzeit festgelegte Normalfrequenz nachträglich veröffentlicht werden. Die Genauigkeit beträgt dabei $\Delta\nu/\nu$ etwa $\pm 1 \cdot 10^{-10}$.

Zur Messung des zeitlichen Ablaufes verschiedenster Vorgänge werden u. a. spezielle Kameras mit extrem hohen Bildfolgen verwendet. Diese werden mittels Drehspiegel auf feststehende Filme erzeugt (man kann einen Film nur bis zu einer bestimmten Grenzgeschwindigkeit entsprechend den Bedingungen in einer mechanischen Kamera bewegen). Mit dem genannten Prinzip erreicht man Bildfolgen um $\approx 10^7$/s und eine Zeitauflösung bis zu $\approx 10^{-9}$ s. Bei Ersatz der mechanisch bewegten Spiegel durch elektronisch-optische Anordnungen (z. B. spezielle Fotokatoden) kann man in geeigneten Fällen noch zwei Ereignisse im Abstand von $\approx 10^{-13}$ s trennen.

Bestimmung sehr großer Zeiträume. Hierbei verwendet man in geeigneten Fällen die Zerfallszeit radioaktiver Substanzen. Das Meßverfahren besteht in einer Konzentrationsbestimmung bestimmter Isotope (zum Teil massenspektrometrisch, vgl. 7.3.2.1.). Eines der wichtigsten Verfahren zur Altersbestimmung von Mineralien ist die *Kalium-Argon-Methode*, die auf dem Zerfall des Isotopes ^{40}K zu ^{40}Ar (10 %) und Calcium (90 %) beruht. Hier muß man das Verhältnis von ^{40}Ar (gasförmig bei Zimmertemperatur) und ^{40}K bestimmen. Von ^{40}K beträgt die Halbwertszeit $T_H = 1{,}26 \times 10^9$ a. Eine weitere Möglichkeit liefert der ^{238}Uraniumzerfall zu Blei ^{206}Pb mit $T_H \approx 4{,}5 \cdot 10^9$ a.

Für organische Substanzen und wesentlich kürzere Zeiten (500 bis 50 000 a) nutzt man die Halbwertszeit des Kohlenstoffes ^{14}C mit $T_H = 5\,730$ a.

Bestimmung sehr kurzer Zeiten ($t \leqq 10^{-11}$ s) bei atomaren oder molekularen Wechselwirkungen erfolgt aus spezifischen Vorgängen in diesen Teilchen.

Die Periodendauer T von Molekülschwingungen der Frequenz ν wird unmittelbar aus den entsprechenden Spektren abgeleitet (über die Wellenlänge λ, wobei gilt $1/T = \nu = c/\lambda$).

Aus vielfältigen spektroskopischen Verfahren gewinnt man die Lebensdauer Δt angeregter Zustände von Teilchen. Aus dem Energiespektrum gewinnt man die natürliche Linienbreite $\Delta\nu$ einer Emission, wobei gilt $\Delta E \cdot \Delta t \geqq h$ (ΔE = Energiebreite, h = Plancksches Wirkungsquantum $= 6{,}626_2 \cdot 10^{-34}$ Ws^2).

7.3.4. Geschwindigkeitsmessung

Auf direktem Weg wird die Geschwindigkeit v auf eine Längen- (oder Winkel-) und eine Zeitmessung zurückgeführt, in SI-Einheiten also in m/s oder rad/s (Winkelgeschwindigkeit) angegeben (km/h ist weiterhin gültig). Bei vorgegebener Strecke dienen zur Auslösung der Zeitmeßeinrichtung mechanische, optische, elektrische, elektronische o. ä. Kontakte. Die Breite von Lichtschranken läßt sich auf wenige Mikrometer reduzieren, so daß diese Meßtechnik in sehr weiten Grenzen einsetzbar ist.

Zur Messung von *Drehzahlen* dienen *Fliehkraftsysteme* (Längsverschiebung von rotierenden Massen auf einer Drehachse mit Übertragung dieser Verschiebung auf ein Hebelsystem oder eine elektrische Anzeige), Schaltkontakte, die durch Nocken auf der Welle betätigt werden bzw. durch magnetische oder kapazitive Geber Impulse auslösen, die einem Zähler zugeführt werden, Stroboskope u. a.

Zu den indirekten Meßverfahren der Geschwindigkeit zählen (als Beispiele) der Doppler-Effekt (Messung einer Differenzfrequenz bzw. Wellenlänge), die Staudruckmessung (vgl. 7.3.8.) (Wasser- oder Luftfahrzeuge), Hitzdrahtanemometer (Widerstandsänderung im Gas- bzw. Luftstrom zur Messung von Strömungsgeschwindigkeiten) oder die Bestimmung der kinetischen Energie E eines Körpers ($v = \sqrt{2E/m}$).

Der *Doppler-Effekt* ist die Verschiebung der Frequenz (bzw. Wellenlänge) einer Strahlung bei einer Relativbewegung v zwischen der Quelle und dem Empfänger (bzw. Beobachter). Man unterscheidet 1. Beobachter ruhend, Quelle bewegt ($v \ll c$); am Ort des Beobachters wird die Frequenz $\nu = \nu_0/(1 \pm v/c)$ wahrgenommen, ν_0 = Frequenz bei $v = 0$). 2. Bei ruhender Quelle und bewegtem Beobachter sieht dieser $\nu = \nu_0(1 \pm v/c)$ mit (+) beide aufeinander zuführend, (−) sich voneinander entfernend. Für eine elektromagnetische Welle ist $c \approx 3 \cdot 10^8$ m/s und für eine Schallwelle ist für c die Schallgeschwindigkeit v_s im betreffenden Medium einzusetzen ($v < v_s$) (Abb. 7.3.4–1).

Bewegt sich das Objekt (mit $v \ll c$), so erscheint die vom Objekt reflektierte Strahlung mit einer Frequenzverschiebung ν_D (gegen die ausgesandte) von $\nu_D = \pm 2v\nu_0/c$. Es sei

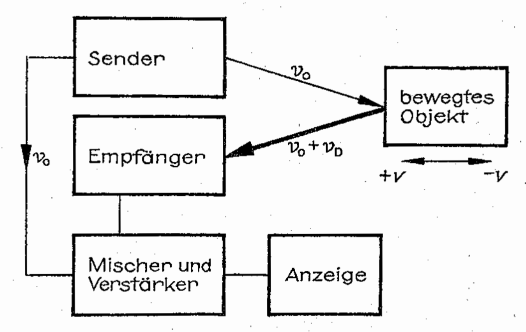


Abb. 7.3.4–1 Prinzipdarstellung des Verkehrsradars

$\nu_0 = 10$ GHz ($\lambda = 3$ cm) und $\nu_D = 10^3$ Hz; daraus $v = c\nu_D/2\nu_0 = 15$ m/s (≈ 54 km/h).
Große Bedeutung hat die Doppler-Verschiebung in der Astronomie zur Bestimmung von Geschwindigkeiten kosmischer Objekte (Umlauf von Doppelsternen, Bewegung von Galaxien u. a.).

7.3.4.1. Schallgeschwindigkeit in Festkörpern kann man u. a. mit einer Anordnung nach Abb. 7.3.4–2 messen. In dem Schallgeber S erzeugt man einen kurzen Impuls (Dauer τ) und mißt dessen Laufzeit t im Probestab L mit dem Empfänger E ($\tau \ll t$). Infolge von Schubspannungen im untersuchten Material sind sowohl Longitudinal- als auch Transversalschwingungen möglich. Lagert man einen dünnen Stab an den Schwingungsknoten, so ergibt sich die Longitudinalgeschwindigkeit v_L aus der Eigenschwingungszeit zu $v_L = 2l \cdot \nu = (E/\varrho)^{1/2}$ (E = Elastizitätsmodul, ϱ = Dichte).

7.3.4.2. Mittlere Geschwindigkeiten von Atomen oder Molekülen kann man aus der betreffenden Gastemperatur zu $\bar{v} = (3kT/m)^{1/2}$ erhalten ($k = 1{,}38 \cdot 10^{-23}$ Ws/K, T = absolute Temperatur [K], m = Masse des Teilchens).

7.3.4.3. Windgeschwindigkeiten mißt man u. a. mit einer Anordnung nach Abb. 7.3.4–3. Die Drehzahl des Flügels ist in weiten Grenzen der Geschwindigkeit proportional (Drehzahlmesser).

7.3.4.4. Geschwindigkeitsänderungen Δv eines Körpers bekannter Masse m können aus der Beschleunigung $b = \mathrm{d}v/\mathrm{d}t$ über die dabei auftretende Kraft $P = mb = \mathrm{d}(mv)/\mathrm{d}t$ ermittelt werden. Die Größe mv heißt Bewegungsgröße oder Impuls. Gegeben sei die Masse des Kraftmessers $m = 1$ kg, die darauf einwirkende Kraft $= 100$ N (vgl. 7.3.5.), das Zeitintervall $\Delta t = 2$ s (Bremszeit), die Endgeschwindigkeit $v_e = 0$. Man erhält eine Anfangsgeschwindigkeit $v_a = b \cdot t = P\Delta t/m = 200$ m/s.
Läßt man z. B. ein Geschoß (Masse m_1) in einen entsprechend aufgehängten Körper der Masse m_2 eindringen und mißt die dabei erfolgende Anhebung Δh des Schwerpunktes, so erhält man die Auftreffgeschwindigkeit (Endgeschwindigkeit $v_e = 0$) $v_a = [(m_1 + m_2)/m_1]\sqrt{2 \cdot g \cdot \Delta h}$, $g \approx 10\,\mathrm{m \cdot s^{-2}}$.
Strömungsgeschwindigkeit vgl. 7.3.8. (Staurohr).

7.3.5. Messung von Kräften

Die Einheit der Kraft ist das Newton (N) $= 1\,\mathrm{kg \cdot m/s^2}$ (abgeleitete SI-Einheit), sie ist über das Gesetz $P = m \cdot b$ an die SI-Einheiten Masse, Länge und Zeit angeschlossen. An der Erdoberfläche wirkt eine Masse von 1 kg auf eine waagerechte Unterlage mit einer (Gravitations-)Kraft von 1 kg $\times\, 9{,}806\,65\,\mathrm{m/s^2} = 9{,}806\,65$ N. Die Richtung dieser Kraft ist bei homogener Schichtung der Erde auf den Erdmittelpunkt hin gerichtet.
Zur Messung einer Kraft dient eine auf eine elastische Längen- oder Querschnittsveränderung bzw. Scherung einer geeigneten Substanz ansprechende Meßanordnung (elastische Verformung). Es gilt $\Delta l/l = P/(q \cdot E)$ mit Δl = durch P bewirkte Längenänderung, l = Länge des Meßstabes, q = dessen Querschnitt und E = Elastizitätsmodul.
Meßwandler für Kräfte sind z. B. Federn (mit bekannter Federkonstante), Kraftmeßdosen nach Abb. 7.3.5–1, Dehnungsmeßstreifen (vgl. 7.3.1.8.). Eine weitere von vielen Möglichkeiten zur Messung von Kräften ist die Anwendung des piezoelektrischen Effektes, insbesondere bei dynamischen Vorgängen. An solchen Substanzen entstehen bei

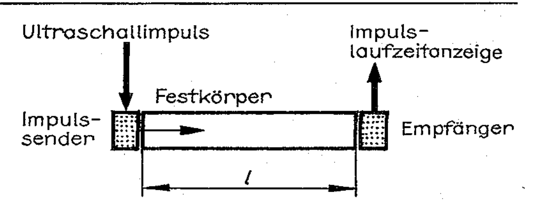


Abb. 7.3.4–2 Messung der Schallgeschwindigkeit v_s im Festkörper (Länge l) mittels Ultraschallimpulse der Laufzeit t; $v_s = l/t$

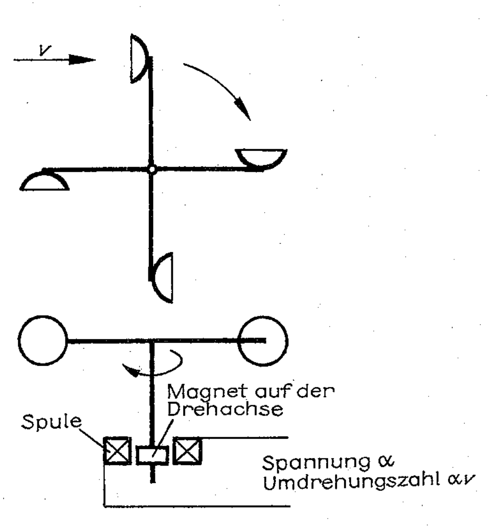


Abb. 7.3.4–3 Messung der Windgeschwindigkeit mit Drehflügel

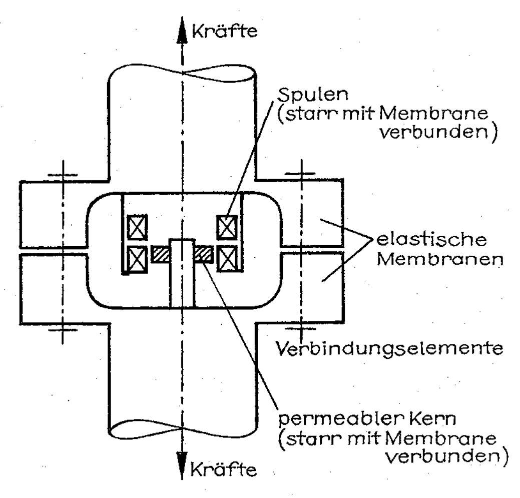


Abb. 7.3.5–1 Induktiver Kraftmesser; entsprechend der Stellung des permeablen Kerns ändert sich die Induktivität der Spulen

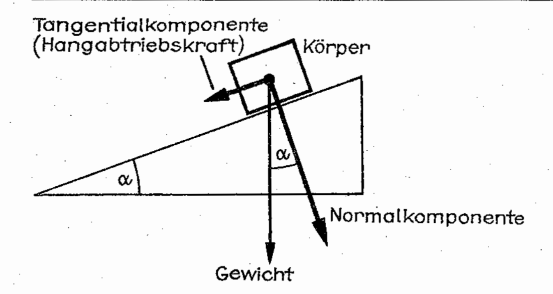


Abb. 7.3.5–2 Schiefe Ebene

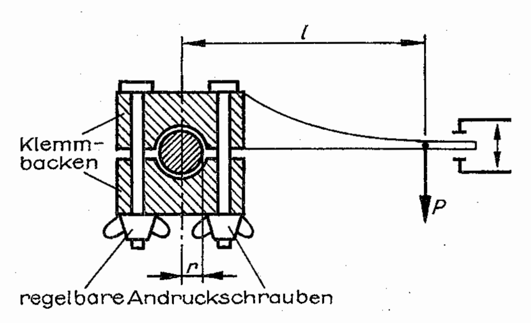


Abb. 7.3.5–3 Gleitende Reibung an einer Welle, Drehmoment $M = P_R \cdot r = P \cdot l$, Reibungskraft P_R am Umfang der Welle mit dem Radius r

Verformungen infolge angreifender Kräfte Spannungsdifferenzen an den Oberflächen. (Bei solchen Substanzen werden bei Deformationen Ladungen erzeugt bzw. verschoben). Im elastischen Bereich besteht dabei die Proportionalität zwischen den wirkenden Kräften P und den entstehenden Spannungen U. Es gilt $P \propto \Delta l$ und $d = \Delta l/U$, wobei Δl = Deformation infolge P bedeutet, d = piezoelektrischer Koeffizient, der für repräsentative Substanzen (Quarz, Bariumtitanat) je nach der Richtung der Deformation zwischen einigen 10^{-10} und 10^{-12} m/V beträgt.
Reibungskräfte treten in vielfältiger Weise auf, z. B. gleitende Reibung oder rollende Reibung. Nach Abb. 7.3.5–2 setze sich der Körper K auf der schiefen Ebene bei dem Winkel $\alpha \geqq \alpha_0 = 22°$ in Bewegung. Dies ist nur möglich, wenn die Tangentialkomponente der angreifenden Kraft $P_s = \tan\alpha \cdot P_n \geqq \mu_0 P_n$, P_n = Normalkomponente der Kraft des Körpers K auf die Unterlage. Der Reibungskoeffizient μ_0 ($= \tan\alpha_0$) beträgt dann $\tan 22° \approx 0{,}4$. Für ein Fahrzeug vom Gewicht $P = 20\,000$ N sei der Koeffizient der rollenden Reibung auf ebener Straße $\mu_0 = 0{,}03$. Zur Fortbewegung auf waagerechter Straße ist eine Kraft $P \geqq 0{,}03 \cdot 20\,000$ N $= 600$ N erforderlich (bei einem Neigungswinkel >arc tan 0,03 ≧ 1,72° würde sich das Fahrzeug von selbst in Bewegung setzen). Entsprechend Abb. 7.3.5–3 ist die Kraft der gleitenden Reibung zwischen der Antriebswelle einer Maschine und den Klemmbacken mit Hilfe der Schrauben einstellbar (variabler Andruck). Die Reibungskraft P_R bildet mit dem Radius r der Welle das *Drehmoment* $P_R \cdot r$. Man stellt nun P_R so ein, daß $P_R \cdot r = P \cdot l$ (P = Kraft, l = Hebellänge), d.h. Gleichgewicht herrscht. Die Länge des Gleitweges ist der Wellenumfang $2\pi r \cdot$ Umdrehung/s $\cdot$ Zeit. Die von der Welle abgegebene Arbeit A beträgt somit $A = 2\pi r v t P_R$ und die Leistung L wird $L = A/t = 2\pi r v P l$. Anstelle der beiden Backen verwendet man häufig eine mit der Welle verbundene Wirbelstrombremse (Metallscheibe zwischen den Polen eines regelbaren Elektromagneten) oder einen belasteten Generator ($L = V \cdot I$). Umgekehrt kann bei bekannter Leistung sofort die Reibungskraft ermittelt werden.
Zur Bestimmung von Kräften über elastische Verformungen können fast alle bereits bei der Längenmessung erwähnten Meßwandler verwendet werden. Kräfte auf Probekörper in Strömungen mißt man über Meßwandler (im einfachsten Fall eine Feder bekannter Federkonstante) gegen einen (festen) Bezugspunkt. Zur Messung der Schubkraft eines Triebwerkes kann man u. a. dieses auf ein entsprechend den Kräften geführtes, aber leicht bewegliches Gestell (z. B. Wagen) montieren und eine Kraftmeßdose (vgl. Abb. 7.3.5–1) zwischen Wagen und einem Widerlager anordnen. Zur direkten Messung von Fliehkräften verwendet man u. a. ebenfalls die erwähnten Meßwandler.

7.3.6. Messung des Elastizitätsmoduls

Die Belastung eines Stabes der Länge l und dem Querschnitt F durch die Kraft P bewirkt eine Längenänderung Δl des Stabes. Innerhalb des Proportionalitätsbereiches zwischen Δl und P gilt $\Delta l = (1/E)\,(lP/F)$, E = Dehnungs- oder Elastizitätsmodul. Häufig wird E bestimmt aus: a) der Längenänderung Δl (vgl. 7.3.1.) eines an einem Ende fest eingespannten Stabes, an dem die (einseitige) Kraft P in Längsrichtung angreift; b) der Durchbiegung Δh eines Stabes unter der Einwirkung der Kraft P nach Abb. 7.3.6–1 (Träger auf zwei Stützen). Bei dieser Messung geht das Flächenträgheitsmoment I_F ein, das von der Querschnittsform abhängig ist. Umgekehrt kann bei bekanntem E des Materials bei unregelmäßigem Querschnitt dessen I_F bestimmt werden (für alle regelmäßigen Querschnitte ist I_F berechenbar).

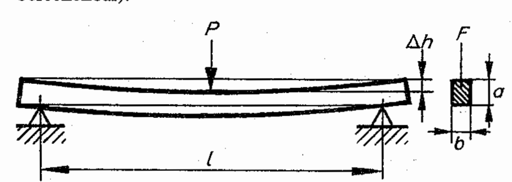


Abb. 7.3.6–1 Elastizitätsmodul E aus Durchbiegung eines Stabes, $E = (1/4)\,(l^3/a^3 \cdot b)\,(P/\Delta h)$, F = Stabquerschnitt = $a \cdot b$

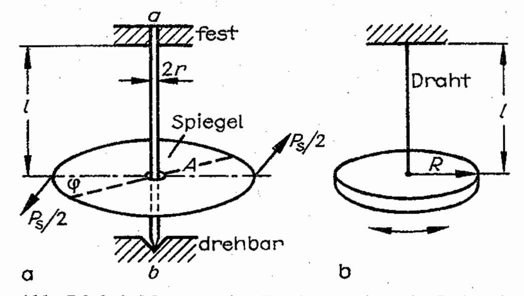


Abb. 7.3.6–2 Messung des Torsionsmoduls G, Stab mit Durchmesser $2r$ der Länge l, P_s = Schubkraft ($\perp l$), Drehmoment $M = R \cdot P_s$, $G = 2l \cdot R \cdot P_s/(\pi \cdot r^4 \cdot \varphi)$

Zum Beispiel ist für den gewählten Stabquerschnitt $I_F = 1/(a^3b)$ (für $a \ll l$) und $E = l^3P/(4a^3b\Delta h)$ (für Baustahl ist $E = 2 \cdot 10^{11}\,N/m^2$).

Der *Torsions-* oder *Schubmodul G* ist ein Maß für die Formelastizität eines Stoffes. Die Messung kann z.B. nach Abb. 7.3.6–2a erfolgen. Ein Stab der Länge l mit kreisförmigem Querschnitt (Radius r und $l \gg r$) ist bei a fest eingespannt und bei b frei drehbar gelagert (Reibung vernachlässigt). Am unteren Stabende können die beiden Kräfte $P_s/2$ am Hebelarm A wirken (Drehmoment $M = P_sA$). Infolgedessen wird der Stab um einen bestimmten (kleinen) Winkel φ verdrillt. Innerhalb der Proportionalitätsgrenze ist $\varphi \propto P_s$. φ kann mittels des mit dem Stab verbundenen Spiegels S bestimmt werden. Man erhält $G = 2lAP_s/(\pi r^4\varphi)$. Aus $A = 0{,}1\,m$, $P_s = 220\,N$, $r = 0{,}01\,m$, $\varphi = 0{,}0175$ (≈ 1 Grad) und $l = 1\,m$ folgt $G \approx 8 \cdot 10^{10}\,N/m^2$ (etwa der Wert für Stahl).

Zur Messung von G an dünnen Drähten eignet sich besser das in Abb. 7.3.6–2b skizzierte Verfahren. Der am oberen Ende fest eingespannte Draht der Länge l trägt am unteren Ende eine Scheibe der Masse m und des Radius R (Massenträgheitsmoment $\Theta_0 = mR^2/2$). Das Ganze versetzt man in Drehschwingungen und erhält G aus der Schwingungszeit $T = 2\pi[\Theta_0 2l/(\pi G r^4)]^{1/2}$ (das Massenträgheitsmoment des Drahtes Θ_1 ist hier vernachlässigt; $\Theta_1 \ll \Theta_0$). Eine für das Strömungs- bzw. Fließverhalten wichtige Größe von Flüssigkeiten und Gasen ist die *Viskosität* η (Zähigkeit). Sie wird angegeben in Pa·s bzw. Ns/m² (abgeleitete SI-Einheit). Man läßt das zu untersuchende Medium durch eine Kapillare strömen, dabei gilt $\eta = \pi r^4(p_1 - p_2) \cdot t/(8Vl)$, $\Delta p = p_1 - p_2$ = Druckdifferenz längs der Kapillare der Länge l, V = in der Zeit t durch die Kapillare geflossenes Volumen des Mediums. Vorausgesetzt ist, daß das Medium an der Kapillarwand haftet (und das Reibungsgesetz gilt); für Δp gilt eine gewisse Beschränkung. Praktisch besteht eine Möglichkeit zur Messung von η in Flüssigkeiten darin, eine Kugel in einem leicht schräg gestellten Rohr mit der zu untersuchenden Substanz durch diese bis zum Boden des Rohres fallen zu lassen und die dazu erforderliche Zeit zu messen. Die Kugel hat nur einen wenig kleineren Durchmesser als das Rohr (Kugel und Rohr erhalten Präzisionsbearbeitung). Die Apparatur muß kalibriert bzw. geeicht werden und es ist eine Thermostatierung erforderlich.

7.3.7. Messung von Oberflächenspannungen

Die Oberflächenspannung γ einer Flüssigkeit kann experimentell direkt als Kraft gemessen werden, die an einer in der Oberfläche liegenden Linie der Länge l senkrecht zu dieser Linie angreift. Nach Abb. 7.3.7–1 besteht eine mögliche Meßanordnung aus einem Drahtbügel mit einem (verschiebbaren) Draht D (etwa 0,1 mm ∅). Der Bügel wird mit der Kraft P (Gewicht an einer dafür geeigneten Waage) langsam aus der Flüssigkeit gezogen, bis die sich unterhalb des Drahtes D gebildete Flüssigkeitslamelle abreißt. Die in diesem Augenblick wirkende Kraft P_1 muß genau bestimmt werden. Unter Vernachlässigung von Korrekturgliedern für den Drahtdurchmesser erhält man als Oberflächenspannung γ der untersuchten Flüssigkeit $\gamma = (P_1 - P_2)/(2l)$. P_2 ist die Kraft beim Herausziehen des Bügels aus der Flüssigkeit ohne den Draht D (Leermessung), z.B. $P_1 - P_2 = 2{,}19 \times 10^{-3}\,N$, $l = 0{,}03\,m$, $\gamma = 7{,}3 \cdot 10^{-2}\,N/m$ (etwa der Wert für Wasser).

Eine andere oft verwendete Methode zur Messung von γ

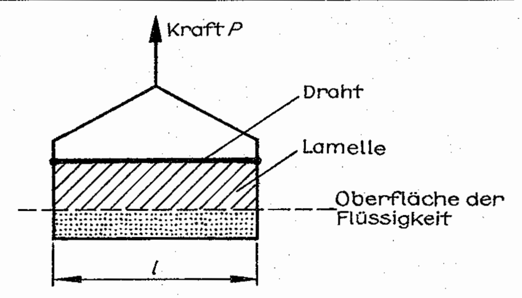


Abb. 7.3.7–1 Messung der Oberflächenspannung $\gamma = P/l$, Draht $\approx 0{,}1\,mm^2$ im Querschnitt, $l \approx 30\,mm$

besteht darin, eine kürzere waagerecht liegende Kapillare vom Radius r mit der zu untersuchenden Flüssigkeit zu füllen und von einer Seite der Kapillare den Druck p so einzustellen, daß sich am freien Ende eine ebene Flüssigkeitsoberfläche ausbildet und man erhält $\gamma = rp/2$. Die Kontrolle der Oberfläche der ebenen Schicht in der Kapillare erfolgt mit optischen Mitteln.

Man kann auch die Steighöhe h einer Flüssigkeit in einer Kapillare messen. Bei vollständiger Benetzung erhält man $\gamma = hr\varrho g/2$, ϱ = Dichte der Flüssigkeit, r = Radius der Kapillare, $g = 9{,}81\,m \cdot s^{-2}$ (Dichte des an die Flüssigkeit angrenzenden Dampfes vernachlässigt); z. B. $r = 10^{-4}\,m$, $\varrho = 10^3\,kg/m^3$, $h = 0{,}15\,m$, $\gamma = 7{,}3 \cdot 10^{-2}\,N/m$.

7.3.8. Druckmessung

In SI-Einheiten erfolgt die Angabe des Druckes in N/m² = Pa (Pascal). Der Normaldruck der Luft, bezogen auf Meereshöhe beträgt $1{,}013 \cdot 10^5$ Pa. Die Druck-Meßverfahren sind sehr vielgestaltig, je nach dem absoluten Druck oder/und der Druckänderung und dem Medium, in oder an dem diese Messungen erfolgen sollen. Zur Druckmessung in Gasen und Flüssigkeiten dienen *Manometer*, von denen einige Arten in der Abb. 7.3.8–1 dargestellt sind. Für Gase sind solche Geräte zwischen etwa 1/1000 und bis zum 1000fachen des Normalluftdruckes verwendbar, für Flüssigkeiten etwa bis zur gleichen oberen Grenze. Die Wirkungsweise beruht auf mechanischen Verspannungen des Meßrohres

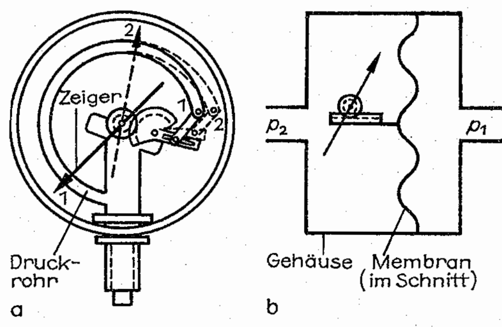


Abb. 7.3.8–1 (a) Röhrenfedermanometer; 1 = Lage des Druckrohres und Zeigers beim Druck p_1, 2 = Lage bei Druck $p_2 > p_1$, (b) Prinzip des Membranmanometers; auch für Differenzdruck $(p_2 - p_1)$, für $p_1 = 0$ folgt Absolutdruck p_2

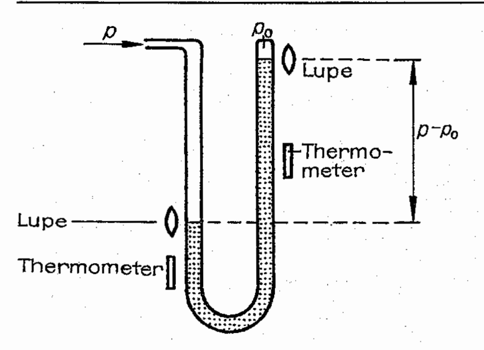


Abb. 7.3.8–2 Stationsbarometer (Flüssigkeitsmanometer), p = zu messender Druck, p_0 = Bezugsdruck (für $p_0 = 0$ wird p absolut gemessen). Die Lupe dient zur Ablesung der Flüssigkeitsoberfläche (längs beider Schenkel befinden sich Skalen mit Noniusablesung)

oder der Meßdose mit einer geeigneten Übertragung auf einen Zeiger. Die Anzeige der Durchbiegung einer Membran kann auch über eine Kapazitätsänderung (oder auch über Lichtzeiger) erfolgen. Solche Differenzdruckmanometer sind bis unter das 10^{-5}fache des Normaldruckes anwendbar, sie sind allgemein gasartunabhängige Absolutdruckmesser. Als Normale gelten Präzisions-Barometer, die die Druckbestimmung in bestimmten Bereichen auf $\Delta p/p < 10^{-4}$ gestatten. Bei diesen Geräten wird die Höhe einer Flüssigkeitssäule sehr genau gemessen, die dem zu bestimmenden Gasdruck das Gleichgewicht hält (Abb. 7.3.8–2). Spezielle Ausführungen (*Kompressionsmanometer*) ermöglichen Messungen bis etwa 10^{-2} Pa.

Nach der Gaskinetik ist der *mittlere Druck p eines Gases* proportional der Anzahl n der Gasmoleküle je Volumeneinheit und deren mittleren kinetischen Energie, d.h. $p = nkT$. Die Proportionalität von n und p (bei T = const.) ist die Grundlage für gasartunabhängige *Wärmeleitfähigkeitsvakuummeter*, bei denen ein elektrisch geheizter Draht je nach der Anzahl der ihn umgebenden Gasteilchen infolge der dadurch unterschiedlichen Wärmeleitung eine bestimmte Temperatur und damit einen bestimmten Widerstand annimmt, der elektrisch angezeigt wird. Der Meßbereich liegt etwa zwischen 10^{-2} und 10^2 Pa. Durch Ionisierung von Gasmolekülen (bzw. Atomen) kann man entsprechend der genannten Proportionalität von p und Teilchenzahl mittels *Ionisations-Vakuummeter* noch Gasdrücke bis unter 10^{-11} Pa messen, indem der durch die Ionen transportierte Strom gemessen wird (1 Ion/s entspricht $1{,}6 \cdot 10^{-19}$ A).

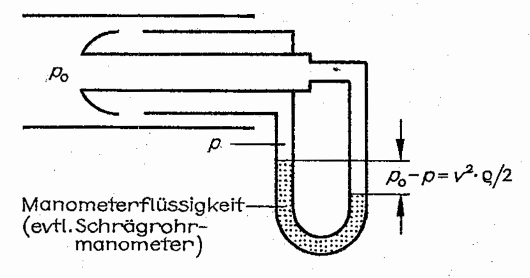


Abb. 7.3.8–3 Staurohr, p_0 = Gesamtdruck, $p_0 - p$ = Staudruck

Zur Messung der *Druckamplituden* von Schallwellen verwendet man u. a. Kondensator-Mikrofone mit einer Empfindlichkeit von $\approx 10^{-3}$ V/0,1 Pa (Richtwert) ± 1 Größenordnung (30 – 10^4 Hz) und einem Meßumfang Δp zwischen $\approx 10^{-5}$ Pa bis 10 Pa (dies entspricht Schallenergien zwischen $\approx 10^{-20}$ und 10^{-8} W/m^2. Die Unterhaltungssprache im Abstand 1 m entspricht etwa einem Schalldruck von 0,1 Pa. Drücke lassen sich, besonders bei dynamischen Prozessen (z. B. Verbrennungskraftmaschinen), u. a. vorteilhaft auch mit piezoelektrischen Meßwandlern messen (vgl. 7.3.5.). In der Hochdruckphysik verwendet man u. a. *Widerstandsmanometer*. Im allgemeinen sinkt der Widerstand von Metallen mit wachsendem allseitigem Druck. Einige dieser Widerstandsänderungen $\Delta R/R$ betragen $\approx 2 \cdot 10^{-6}$ je 10^5 N/m^2. Manganin zeigt entgegengesetztes Verhalten, aber eine größere Abhängigkeit des Widerstandes vom Druck und wird deshalb bevorzugt zu Hochdruckmessungen verwendet. Für die Strömung idealer Flüssigkeiten (inkompressibel) und für Strömungsgeschwindigkeiten v < Schallgeschwindigkeit v_s (im betreffenden Medium) gilt $Q = F_1 v_1 = F_2 v_2$, Q = Durchflußmenge, F_1, F_2 = verschiedene Querschnitte innerhalb des betrachteten Gebietes. In Gebieten größerer Geschwindigkeit muß verminderter Druck herrschen. Es ist $p + \varrho \cdot v^2/2 = p_0$ (= const.), p_0 = Gesamtdruck, p = statischer Druck, $\varrho v^2/2$ = Staudruck, ϱ = Dichte des Mediums. Die Meßanordnung ist in Abb. 7.3.8–3 skizziert. Mittels des Staurohres kann man die Geschwindigkeit von Wasser- oder Luftfahrzeugen messen. Bei gegebenen Querschnitten F_1 und F_2 und ϱ erhält man aus $\Delta p = p_1 - p_2 = \varrho(v_2^2 - v_1^2)/2$ die Gesamtmenge Q zu $Q = F_2 v_2 = F_2[2\Delta p/(\varrho(1 - (F_1/F_2)^2))]^{1/2}$ (Durchflußmesser).

Den Luftdruck p in der Atmosphäre ermittelt man nach $p = p_0 \exp(-h/H_0)$, wobei H_0 die Höhe einer gleichmäßigen Atmosphäre ist (T = const.). Für einen Temperaturgradienten von 6,5°/100 m und $H \leqq 11{,}5$ km folgt die Druckverteilung in der Atmosphäre zu $p_H/p_0 = (1 - 6{,}5\,H/288)^{5{,}255}$ mit H = Höhe in km und p_0 = Druck in der Höhe $H = 0$.

7.3.9. Temperaturmessung

7.3.9.1. Temperaturskale. Die Maßeinheit der *thermodynamischen Temperatur* ist das *Kelvin* (K). Die thermodynamische Temperaturskale beginnt am absoluten Nullpunkt mit $T = 0$ K. Für die Festlegung der Maßeinheit ordnet man durch Definition dem Tripelpunkt von reinem Wasser natürlicher Isotopenzusammensetzung (vgl. 9.1.3.1.) den genauen Temperaturwert $T_{tr} = 273{,}16$ K zu, so daß 1 K = T_{tr}(Wasser)/273,16 gilt. Mit dieser Definition verfügt man jedoch noch nicht über eine praktisch brauchbare Vorschrift für die Eichung von Temperaturmeßgeräten. Diese Aufgabe wird durch die sogenannte *Internationale Praktische Temperaturskale von 1968* (IPTS-68) für den Temperaturbereich oberhalb 13,8 K und durch die sogenannte *Provisorische Temperaturskale von 1976* (EPT-76) für den Temperaturbereich von 0,5 bis 30 K gelöst. Die IPTS-68 basiert auf einer Reihe von Temperaturen für Tripelpunkte, Siedepunkte und Erstarrungspunkte ausgewählter reiner Stoffe, den sogenannten *definierenden Fixpunkten* (Tab. 7.3.9–1). Die Temperaturen zwischen den definierenden Fixpunkten werden durch Interpolation mit dem Platin-Widerstandsthermometer (vgl. 7.3.9.5.) im Bereich T = 13,81 bis 903,905 K und dem Platin/Platin-Rhodium-Thermoelement (vgl. 7.3.9.6.) im Bereich T = 903,905 bis 1337,58 K ermittelt. Die Extrapolation über die Erstarrungstemperatur

des Goldes hinaus erfolgt mit einem Strahlungsthermometer (vgl. 7.3.9.7.). Um die Interpolation zwischen den definierenden Fixpunkten und die Extrapolation in den Hochtemperaturbereich zu erleichtern, enthält die IPTS-68 noch ≈40 sogenannte *sekundäre Bezugspunkte* zwischen 24,56 K (Tripelpunkt des Neons) und 3 695 K (Schmelzpunkt des Wolframs).

Ebenso wie die IPTS basiert auch die EPT-76 auf einer Reihe von Fixpunkten, die in Tab. 7.3.9–2 zusammengestellt sind. Neben den Siede- und Tripelpunkten geeigneter Gase werden hier die Supraleitungssprungpunkte (definiert als die Temperatur des Übergangs zwischen dem supraleitenden und normalleitenden Zustand bei Abwesenheit eines Magnetfeldes) einiger Reinstmetalle als Fixpunkte verwendet. Bezüglich der für die Temperaturbereiche zwischen den Fixpunkten und für Temperaturen unterhalb 0,5 K zu verwendenden Interpolations- bzw. Extrapolationsinstrumente enthält die EPT-76 keine Vorschriften. Meist werden hierfür Germanium- oder Kohle-Widerstandsthermometer (vgl. 7.3.9.5.) oder das Helium-Dampfdruckthermometer (vgl. 7.3.9.4.) eingesetzt.

Neben der thermodynamischen Temperaturskala wird auch die *Celsius-Skale* für Temperaturangaben verwendet. Die Celsius-Temperatur ϑ ist definiert als die Differenz aus der thermodynamischen Temperatur T und der Temperatur $T_0 = 273{,}15$ K des Eispunkts von Wasser: $\vartheta = T - T_0$. Bei der Angabe von Celsius-Temperaturen wird für die Benennung der Temperatureinheit nicht das Kelvin (K), sondern das *Grad Celsius* (°C) verwendet.

7.3.9.2. Gasthermometer. In den Gasthermometern wird der durch die Zustandsgleichung von Gasen gegebene Zusammenhang zwischen dem Druck p, dem Volumen V und der Temperatur T eines Gases für die Temperaturmessung ausgenutzt. Für ein ideales Gas, in dem die Gasteilchen keinerlei anziehende oder abstoßende Kräfte aufeinander ausüben, lautet diese Zustandsgleichung $pV = nRT$, wobei n die im Volumen V enthaltene Stoffmenge in Mol und R die molare Gaskonstante ist. Unter Ausnutzung dieser Gleichung läßt sich die Temperaturmessung auf eine Druck- und Volumenmessung (vgl. 7.3.1., 7.3.8.) zurückführen. Die wichtigsten Teile eines Gasthermometers sind schematisch in Abb. 7.3.9–3 dargestellt. Das Gasthermometergefäß mit dem Volumen V befindet sich auf der zu messenden Temperatur T. Es ist über kapillare Leitungen mit dem durch Heben oder Senken des Quecksilberspiegels meßbar veränderlichen Volumen V_z (bis zur Marke c) und dem hier als U-Rohr-Manometer (vgl. 7.3.8.) dargestellten Druckmeßgerät verbunden, in dem der Quecksilberstand im linken Schenkel stets so einzustellen ist, daß er bis zur Marke d reicht. Die Regulierung des Quecksilberstandes zur Einstellung von V_z und im Manometer erfolgt über die Ventile 1 bzw. 2. Bei der Messung befinden sich alle Teile des Gasthermometers oberhalb der Marke b auf einer konstanten Bezugstemperatur T_b. Meist befindet sich dieser Teil des Gasthermometers in einem Thermostaten (vgl. 7.1.1.), wobei für T_b Werte um 300 K üblich sind. Das auf T_b befindliche Gasvolumen besteht aus dem variablen Volumen V_z und dem durch die Kapillarleitungen und den quecksilberfreien Teil des linken Schenkels des U-Rohr-Manometers konstanten Volumen V_b. Im Volumen V_k des Kapillarstückes zwischen den Marken a und b wird sich ein Temperaturgefälle ausbilden. Da V_k jedoch gegenüber den anderen Volumina klein ist, wird für die in V_k befindliche Gasmenge angenommen, daß sie sich auf einer mittleren Temperatur $T_k = (T + T_b)/2$ befindet.

Bei jeder gasthermometrischen Messung werden Druck und Volumen der im Gasthermometer eingeschlossenen und konstanten Gasmasse in zwei unterschiedlichen Zuständen 1 und 2 gemessen. Da die gesamte Stoffmenge, d.h. die Summe aus den Stoffmengen n_b in V_b und V_z, n_k in V_k und n in V konstant bleibt, gilt

$$(n_b + n_k + n)_{\text{Zustand 1}} = (n_b + n_k + n)_{\text{Zustand 2}}.$$

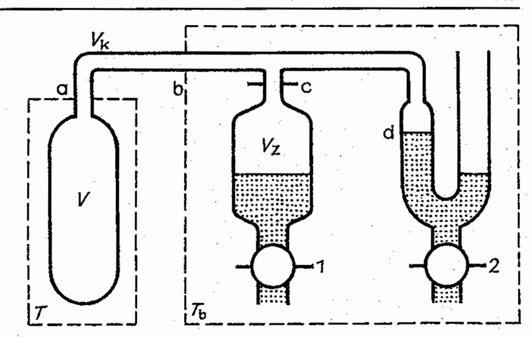


Abb. 7.3.9–3 Schema eines Gasthermometers

Tab. 7.3.9–1 Definierende Fixpunkte der IPTS-68. Sofern nicht anders angegeben, beziehen sich die Temperaturangaben für Siede- und Erstarrungspunkte auf einen Druck von $p_0 = 101\,325$ Pa (normaler Atmosphärendruck)

Zustand	T in K
Tripelpunkt des Wasserstoffs	13,81
Siedepunkte des Wasserstoffs bei $p = 33\,330{,}6$ Pa	17,042
Siedepunkt des Wasserstoffs	20,28
Siedepunkt des Neons	27,102
Tripelpunkt des Sauerstoffs	54,361
Tripelpunkt des Argons	83,798
Siedepunkt des Sauerstoffs	90,188
Tripelpunkt des Wassers	273,16
Siedepunkt des Wassers	373,15
Erstarrungspunkt des Zinns	505,1181
Erstarrungspunkt des Zinks	692,73
Erstarrungspunkt des Silbers	1235,08
Erstarrungspunkt des Goldes	1337,58

Tab. 7.3.9–2 Fixpunkte der EPT-76

Fixpunkt	T in K
Supraleitungssprungpunkt des Cadmiums	0,519
Supraleitungssprungpunkt des Zinks	0,851
Supraleitungssprungpunkt des Aluminiums	1,1796
Supraleitungssprungpunkt des Indiums	3,4145
Siedepunkt des Heliums (^{4}He)	4,2221
Supraleitungssprungpunkt des Bleis	7,1999
Tripelpunkt des Wasserstoffs	13,8044
Siedepunkt des Wasserstoffs bei $p = 33\,330{,}6$ Pa	17,0373
Siedepunkt des Wasserstoffs	20,2734
Tripelpunkt des Neons	24,5591
Siedepunkt des Neons	27,102

Diese Gleichung bildet zusammen mit der Zustandsgleichung, die n_b, n_k und n mit den entsprechenden Volumina, Drücken und Temperaturen verknüpft, die Grundlage der Gasthermometrie. Entspricht dem Zustand 1 ein Druck p_1, ein variables Volumen V_{z1}, eine Temperatur $T = T_1$ und eine mittlere Temperatur T_{k1} in V_k und dem Zustand 2 ein anderer Wertesatz p_2, V_{z2}, $T = T_2$ und T_{k2}, so ergibt sich für die Grundgleichung des Gasthermometers die Beziehung

$$p_1\left(\frac{V}{T_1} + \frac{V_b + V_{z1}}{T_b} + \frac{V_k}{T_{k1}}\right) = p_2\left(\frac{V}{T_2} + \frac{V_b + V_{z2}}{T_b} + \frac{V_k}{T_{k2}}\right).$$

Die Größen V_b, V_k, V und T_b sind durch die Konstruktion des Gasthermometers vorgegeben. Nimmt man noch an, daß T_1 bekannt ist (meist wählt man $T_1 = T_b$), so erhält man die gesuchte Temperatur T_2 aus der direkten Messung von p_1, p_2, V_{z1} und V_{z2}. In der praktischen Temperaturmessung mit Gasthermometern wählt man nun nicht die durch die obige Gleichung beschriebene allgemeinste Zustandsänderung der eingeschlossenen Gasmenge, sondern hält entweder das variable Volumen V_z (Methode konstanten Volumens), den Druck p (Methode konstanten Drucks) oder die Temperatur T (Methode konstanter Temperatur) während der Zustandsänderung konstant. Bei der am häufigsten verwendeten Methode konstanten Volumens wählt man $T_1 = T_b$ und $V_{z1} = V_{z2} = 0$, so daß die Bestimmung der unbekannten Temperatur durch zwei Druckmessungen erfolgt. Bei der Methode konstanten Drucks wählt man wieder $T_1 = T_b$ und mißt die für die Einstellung gleicher Drücke $p_1 = p_2$ erforderliche Volumenänderung $V_{z1} - V_{z2}$. Im Falle der Methode konstanter Temperatur schließlich ist $T_1 = T_2$. Diese Temperatur läßt sich aus der Grundgleichung des Gasthermometers ermitteln, indem man die Drücke p_1 und p_2 bei zwei unterschiedlichen variablen Volumina V_{z1} und V_{z2} mißt, wobei zur Vereinfachung der Messung noch $V_{z1} = 0$ gewählt werden kann.

Gasthermometer können für Temperaturmessungen im Bereich $T = 1$ bis 1900 K eingesetzt werden. Als Materialien für die Herstellung des Thermometergefäßes kommen für den Temperaturbereich unterhalb ≈ 1300 K vor allem Quarzglas, für Temperaturen oberhalb 1300 K Platinlegierungen in Frage. Die wichtigsten Füllgase sind Helium, Argon und Stickstoff.

Wird eine hohe Genauigkeit in der Temperaturmessung angestrebt, so ist die durch Wärmeausdehnung bedingte Änderung des Volumens V des Temperaturfühlers (vgl. Abb. 7.3.9–3) mit der Temperatur zu berücksichtigen (vgl. 7.3.11.2.). Außerdem können insbesondere bei tiefen Temperaturen Abweichungen der tatsächlichen Zustandsgleichung des verwendeten Füllgases von der Zustandsgleichung des idealen Gases eine Rolle spielen.

7.3.9.3. Ausdehnungsthermometer. Alle Flüssigkeiten und festen Körper zeigen eine mehr oder weniger ausgeprägte Änderung ihres Volumens mit der Temperatur (vgl. 7.3.11.2.). Diese Volumenänderung wird durch die Beziehung $V(T) = V(T_0)\,[1 + \beta(T - T_0)]$ beschrieben, wobei der *kubische Ausdehnungskoeffizient* β selbst noch von der Temperatur abhängen kann.

Flüssigkeitsthermometer nutzen diesen Effekt aus. Sie bestehen aus einem Gefäß mit dem Volumen V_0, das vollständig mit der Thermometerflüssigkeit gefüllt ist, und einer sich daran anschließenden Kapillaren aus dem gleichen Material, deren Füllung mit Thermometerflüssigkeit von der Temperatur abhängt. Betrachtet man zwei Temperaturen T_0 und $T > T_0$, die einer Füllhöhe h_0 und h der Kapillaren mit Thermometerflüssigkeit entsprechen, so ergibt sich für die Volumenänderung der Flüssigkeit beim Übergang von T_0 zu T die Relation

$$V(T) - V(T_0) = V(T_0)\beta(T - T_0) = q(h - h_0),$$

wenn q der innere Querschnitt der Kapillaren ist. Die Bestimmung von T aus der gemessenen Füllhöhendifferenz $h - h_0$ erfordert somit die Kenntnis von T_0. In der Regel wird für T_0 der Nullpunkt der Celsius-Skale, $T_0 = 273{,}15$ K, gewählt (vgl. 7.3.9.1.), so daß sich die einfache Beziehung $\vartheta = (h - h_0)q/\beta V(0\,°C)$ ergibt.

Als Materialien für die Herstellung der Thermometergefäße werden spezielle Gläser und Quarzglas verwendet. Die am häufigsten eingesetzten Thermometerflüssigkeiten sind Pentan (für den Temperaturbereich $T = 85$ bis 309 K), Ethanol ($T = 160$ bis 350 K) und Quecksilber ($T = 235$ bis 1050 K).

Zur Erleichterung der Temperaturbestimmung wird an der Kapillaren eine in Grad Celsius geeichte Skale angebracht.

Metallausdehnungsthermometer beruhen auf der Tatsache, daß feste Körper eine vorgegebene geometrische Form haben und die Änderung ihrer geometrischen Abmessung in einer bestimmten Richtung der Beziehung $l(T) = l(T_0)\,[1 + \alpha_l(T - T_0)]$ folgt. Die Größe α_l wird als *linearer thermischer Ausdehnungskoeffizient* bezeichnet. Metallausdehnungsthermometer werden in zwei Ausführungsformen verwendet. Beim *Stabthermometer* besteht der Temperaturfühler aus einem Metallrohr mit hohem linearem Ausdehnungskoeffizienten α_{l1} (Aluminium, Messing, Nickel, Chrom-Nickel-Legierungen, Stahl) und einem mit seiner Bodenplatte fest verbundenen, im Inneren des Rohres befindlichen Stab aus einem Material mit geringerem linearem Ausdehnungskoeffizienten α_{l2} (Quarz, Porzellan, Invar). Haben bei einer Temperatur T_0 der Stab und das Metallrohr die gleiche Länge $l(T_0)$, so wird bei einer Temperatur $T \neq T_0$ eine Längendifferenz zwischen Stab und Rohr auftreten, die sich aus der allgemeinen Formel für $l(T)$ zu $\Delta l = l(T_0)\,[\alpha_{l1} - \alpha_{l2}]\,[T - T_0]$ ergibt. Die Messung von Δl (vgl. 7.3.1.) erlaubt dann bei Kenntnis von T_0 die Bestimmung von T.

Bimetall-Thermometer. Hier werden zwei Metallstreifen mit möglichst großem Unterschied in ihren linearen Ausdehnungskoeffizienten fest miteinander verlötet oder aufeinander gewalzt (Abb. 7.3.9–4). Wird jetzt das eine Ende des geradlinigen Bimetallstreifens der Länge l und der Dicke d festgehalten und die Temperatur um ΔT geändert, so wird sich das andere Ende des Streifens aufgrund der unterschiedlichen thermischen Ausdehnungskoeffizienten der beiden Metalle senkrecht zur Streifenrichtung um einen Betrag Δh verschieben, der sich zu $\Delta h = al^2d^{-1}\Delta T$ ergibt. Die Proportionalitätskonstante a hängt von den linearen thermischen Ausdehnungskoeffizienten und den elastischen Konstanten der beiden Metalle ab und wird experimentell durch Kalibrierung bestimmt. Um die Empfind-

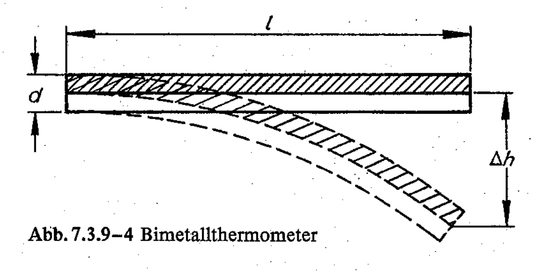


Abb. 7.3.9–4 Bimetallthermometer

lichkeit des Bimetallthermometers zu erhöhen, wird die Länge l dadurch vergrößert, indem man den Bimetallstreifen zu einer Spirale aufwickelt. Bimetallthermometer zeigen eine geringe thermische Trägheit und werden deshalb häufig für meteorologische Zwecke (z.B. Messung von Lufttemperaturen in Registriergeräten) eingesetzt.

7.3.9.4. Dampfdruckthermometer. Schließt man einen reinen flüssigen Stoff in einem evakuierten Gefäß ein, so wird sich in Abhängigkeit von der Temperatur über der Flüssigkeitsoberfläche ein Sättigungsdampfdruck p einstellen, der dem Gleichgewicht zwischen flüssiger und gasförmiger Phase dieses Stoffes entspricht. Die Abhängigkeit des Dampfdruckes von der Temperatur wird durch die Clausius-Clapeyron-Gleichung beschrieben (vgl. 9.1.3.1.) und folgt in vielen Fällen der einfachen Beziehung $p = a \exp(-b/T)$, wobei a und b stoffspezifische Parameter sind. Diese Relation wird in den Dampfdruckthermometern zur Temperaturmessung ausgenutzt. Der Meßfühler des Dampfdruckthermometers besteht aus einer Metallkammer, die teilweise mit der verwendeten Flüssigkeit gefüllt ist. Diese Meßkammer ist über eine Kapillarleitung mit einem Druckmeßgerät verbunden (U-Rohr-Manometer bei Drükken bis zu $\approx 1{,}3 \cdot 10^5$ Pa, Röhrenfedermanometer bei höheren Drücken vgl. 7.3.8.).

Dampfdruckthermometer mit verflüssigten Gasen als Meßflüssigkeiten (Edelgase, Wasserstoff, Sauerstoff, Stickstoff) sind für die Tieftemperaturmessung wichtig. Mit dem Helium-Dampfdruckthermometer können genaue Temperaturmessungen im Bereich $T = 0{,}2$ bis $5{,}2$ K ausgeführt werden. Für technische Temperaturmessungen sind Dampfdruckthermometer für den Temperaturbereich $T = 230$ bis 630 K entwickelt worden. Hier werden leicht verdampfende organische Flüssigkeiten (z. B. Propan, Pentan, Hexan, Xylol) als Meßflüssigkeiten verwendet.

7.3.9.5. Widerstandsthermometer nutzen die Temperaturabhängigkeit des elektrischen Widerstandes von Metallen und Halbleitern für die Temperaturmessung aus. In Metallen wird zur Beschreibung der Temperaturabhängigkeit des Widerstandes R die Beziehung $R(T) = R(T_0)\,[1 + \beta_R(T - T_0)]$ verwendet, wobei der *Temperaturkoeffizient des elektrischen Widerstandes* β_R selbst noch von der Temperatur abhängt. Als Bezugstemperatur T_0 wird meist der Eispunkt des Wassers (Nullpunkt der Celsius-Skale), $T_0 = 273{,}15$ K, gewählt. Platin ist das für Widerstandsthermometer am häufigsten verwendete Metall, da es gegenüber anderen Metallen eine Reihe von Vorteilen aufweist. Es kann sehr rein hergestellt werden, hat einen hohen Schmelzpunkt (2 042 K) und ist auch bei hohen Temperaturen noch sehr korrosionsbeständig.

Platinwiderstandsthermometer. Dafür werden Platindrähte mit Durchmessern zwischen 0,05 und 0,6 mm verwendet. Diese Drähte werden möglichst spannungsfrei und bifilar auf einen Träger gewickelt. Je nach Temperaturbereich werden Träger aus Glas, Glimmer, keramischen Massen und für hohe Temperaturen auch aus synthetischem Saphir verwendet. Diese Widerstandswicklungen werden in Schutzrohren aus Glas, Quarzglas, keramischen Stoffen oder Platin, im Falle technischer Thermometer auch aus Nickel oder nichtrostendem Stahl untergebracht. Zur Vermeidung von Korrosionseffekten werden die Schutzrohre meist mit einer Gasfüllung (Helium, Argon, Stickstoff, trockene Luft) versehen und dann gasdicht abgeschlossen. Der Einsatzbereich des Platinwiderstandsthermometers ist nach oben durch den Schmelzpunkt des Platins (2 042 K) begrenzt. Die untere Einsatzgrenze liegt bei ≈ 2 K und ist durch den Abfall des Temperaturkoeffizienten β_R des elektrischen Widerstands unterhalb ≈ 90 K bedingt (Abb. 7.3.9–5).

Geeignet hergestellte Halbleiter zeigen in bestimmten Temperaturbereichen eine exponentielle Abhängigkeit des elektrischen Widerstandes von der Temperatur, $R(T) = A \exp(B/T)$, wobei A und B selbst noch schwach von der Temperatur abhängige materialspezifische und durch das Herstellungsverfahren modifizierbare Parameter sind. Im Bereich tiefer Temperaturen werden Germaniumeinkristalle und aus Kohleschichten bestehende Widerstände für die Temperaturmessung eingesetzt.

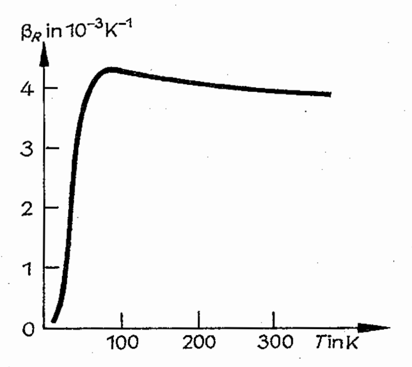


Abb. 7.3.9–5 Temperaturkoeffizient β_R des elektrischen Widerstandes von Platin

Germaniumwiderstandsthermometer. Bei ihnen wird der Einkristall spannungsfrei in ein aus einem Edelmetall bestehendes Schutzrohr eingebaut, das mit einem Schutzgas gefüllt und dann gasdicht abgeschlossen wird. Das Einsatzgebiet liegt im Temperaturbereich $T = 1$ bis 20 K.

Kohlewiderstandsthermometer bestehen aus auf isolierende Träger aufgedampften Kohleschichten und erlauben Temperaturmessungen im Bereich $T = 0{,}001$ K bis 50 K.

Thermistoren sind Halbleiter-Widerstandsthermometer, die aus gesinterten Metalloxiden (Fe_3O_4, Zn_2TiO_4, $MgCr_2O_4$) bestehen und in handelsüblichen Ausführungsformen für Temperaturmessungen im Bereich $T = 200$ bis 600 K ausgelegt sind.

Neben Halbleiterwiderständen werden auch Halbleiterdioden und Transistoren als Temperaturfühler verwendet. Dabei wird die exponentielle Abhängigkeit der Strom-Spannungs-Kennlinien dieser Bauelemente von der Temperatur ausgenutzt. Mit Siliciumdioden lassen sich Temperaturmessungen im Bereich $T = 1{,}5$ bis 400 K ausführen, für Dioden auf der Basis von Galliumarsenid sind Messungen im Bereich $T = 0{,}3$ bis 400 K möglich. Für die Temperaturmessung wird dabei die Temperaturabhängigkeit der Durchlaßspannung der Diode für einen konstanten Durchlaßstrom ausgenutzt.

7.3.9.6. Thermoelemente. Die Thermoelemente oder Thermopaare basieren auf dem *Seebeck-Effekt*. Werden die Drähte zweier unterschiedlicher Metalle a und b entsprechend dem in Abb. 7.3.9–6 gezeigten Schema an zwei Stellen miteinander verschweißt oder verlötet und bringt man diese beiden Verbindungsstellen auf unterschiedliche Temperaturen T_0 und T, so tritt zwischen den Leiterenden 1 und 2 eine elektrische Spannung auf, die bei $T = T_0$ gleich Null wird. Diese Spannung wird als Thermospannung $E_{ab}(T, T_0)$ bezeichnet, sie hängt von der Art der verwendeten Metalle

und von den Temperaturen T und T_0 ab. Alle Thermoelemente haben den in Abb. 7.3.9–6 gezeigten einfachen Aufbau. Es werden Drähte mit Durchmessern zwischen 0,1 und 1 mm verwendet, die Drahtlängen richten sich nach dem konkreten Anwendungszweck. Die beiden unterschiedlichen Metalle, die für die Herstellung eines Thermoelements erforderlich sind, werden so gewählt, daß sie in dem bei der Messung interessierenden Temperaturbereich nicht nur eine möglichst große Thermospannung, sondern auch eine möglichst große Änderung der Thermospannung mit der Temperatur aufweisen. Diese Größe wird als *differentielle Thermospannung* oder *Seebeck-Koeffizient* S_{ab} bezeichnet und ist durch $S_{ab}(T) = \mathrm{d}E_{ab}(T, T_0)/\mathrm{d}T$ definiert. In vielen praktisch verwendeten Thermoelementen ist $E_{ab}(T, T_0)$ eine angenähert lineare Funktion der Temperaturdifferenz $T - T_0$. Als Bezugstemperatur T_0 wird in der Regel der Eispunkt des Wassers gewählt, $T_0 = 273{,}15$ K.

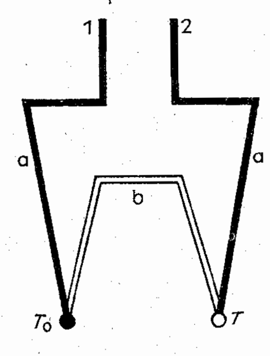


Abb. 7.3.9–6 Aufbau eines Thermoelementes

Gebräuchliche Paare von Metallen oder Metallegierungen, die in Thermoelementen für mittlere und hohe Temperaturen eingesetzt werden, sind in Tab. 7.3.9–7 zusammengestellt. Für Tieftemperaturmessungen sind Thermoelemente entwickelt worden, die aus Kupfer oder Chromel einerseits und Gold oder Silber mit geringfügigen Beimengungen von Eisen oder Cobalt andererseits bestehen. Das Thermoelement Chromel – Gold/0,02 % Eisen ist für Messungen bis herunter zu 0,5 K geeignet.

7.3.9.7. Strahlungsthermometer (Pyrometer). Unter der *Temperaturstrahlung* oder *Wärmestrahlung* eines Körpers versteht man die von seiner Oberfläche ausgehende elektromagnetische Strahlung, deren Eigenschaften von der Temperatur des Körpers und der physikalisch-chemischen Beschaffenheit der strahlenden Oberfläche abhängen (vgl. 7.5.). Durch Messung dieser Strahlung kann die Temperatur des strahlenden Körpers bestimmt werden, indem man die die Temperaturstrahlung beschreibenden sogenannten Strahlungsgesetze ausnutzt. Man unterscheidet zwischen visuellen und objektiven Pyrometern. Bei den *visuellen Pyrometern* erfolgt die Bewertung der Temperaturstrahlung mit dem menschlichen Auge, der weitaus wichtigste Vertreter dieser Klasse von Pyrometern ist das *Glühfadenpyrometer*. Sein prinzipieller Aufbau ist in Abb. 7.3.9–8 gezeigt. Das strahlende Meßobjekt wird mit einem Fernrohr betrachtet, dessen Objektiv in der Mittelebene der Pyrometerlampe ein reelles Bild der strahlenden Oberfläche des Objekts erzeugt. Das Auge betrachtet durch ein Okular und ein Farbfilter sowohl dieses reelle Bild als auch die Spitze des Glühfadens der Pyrometerlampe und vergleicht dabei die Helligkeit des Bildes mit der Helligkeit des Glühfadens. Schließlich wird durch elektrische Helligkeitsregelung des Glühfadens der Pyrometerlampe und, falls das nicht ausreicht, durch meßbare Abschwächung der Strahlung des Meßobjekts mit Hilfe des hinter dem Objektiv angeordneten Schwächungsfilters auf gleiche Helligkeit des Bildes des Meßobjektes und der Spitze des Glühfadens eingestellt. Aus dem hierfür erforderlichen elektrischen Strom durch die Pyrometerlampe, der die Temperatur der Glühfadenspitze bestimmt, kann die Temperatur des Meßobjektes ermittelt werden. Als Pyrometerlampen werden meist Wolframbandlampen eingesetzt, die Betriebstemperaturen bis zu $\approx 2\,800$ K zulassen. Als Farbfilter dienen eingefärbte Glasfilter oder Interferenzfilter, deren Durchlässigkeitsmaximum im roten Spektralbereich (Wellenlänge ≈ 650 nm) liegt. Der Einsatzbereich des Glühfadenpyrometers wie auch anderer visueller Pyrometer ist durch die Bedingung festgelegt, daß die Temperatur des Meßobjektes so hoch sein muß, daß die abgegebene Strahlung noch einen durch das Auge wahrnehmbaren Anteil an Strahlungsleistung im sichtbaren Spektralbereich enthalten muß. Diese Bedingung ist für Temperaturen oberhalb ≈ 950 K erfüllt.

Tab. 7.3.9–7 Thermoelemente für mittlere und hohe Temperaturen

Metallpaar	Temperaturbereich in K
Kupfer – Konstantan[1]	2–700
Eisen – Konstantan[1]	70–1 300
Chromel[2] – Alumel[3]	220–1 600
Platin – Platin/10 % Rhodium	250–2 100
Iridium – Iridium/60 % Rhodium	270–2 300
Wolfram – Molybdän	1 300–2 800
Wolfram – Wolfram/26 % Rhenium	300–3 100

[1] Konstantan: 55 % Kupfer + 45 % Nickel
[2] Chromel: 89 % Nickel + 10 % Chromium + 1 % Eisen
[3] Alumel: 94 % Nickel + 2 % Aluminium + 1,5 % Silicium + 2,5 % Mangan

Objektive Pyrometer verwenden physikalische Strahlungsempfänger (vgl. 7.5.) für die Bewertung der vom Meßobjekt emittierten Strahlung. Man unterscheidet drei Arten von objektiven Pyrometern.

Beim *Gesamtstrahlungspyrometer* wird die Strahlungsleistung des sich auf der Temperatur T befindlichen Meßobjektes und die Strahlungsleistung eines sich auf einer bekannten Temperatur T_0 befindlichen Vergleichsstrahlers gemessen. Aus dem Verhältnis der beiden Strahlungsleistungen läßt sich dann mit Hilfe des Stefan-Boltzmann-Gesetzes (vgl. 7.5.) die Temperatur T bestimmen. Als Strahlungsempfänger werden in diesem Falle Strahlungsthermoelemente oder Bolometer (vgl. 7.5.) eingesetzt.

Das *Strahldichtepyrometer* unterscheidet sich vom Gesamtstrahlungspyrometer nur dadurch, daß nicht die Gesamtstrahlungsleistung, sondern nur die Strahlungsleistung des Meßobjektes und des Vergleichsstrahlers bei einer bestimmten Wellenlänge gemessen werden. Das kann in einfacher Weise dadurch geschehen, daß man in einem Gesamtstrahlungspyrometer vor dem Strahlungsempfänger ein Farbfilter anordnet. Beim Strahldichtepyrometer ergibt sich die gesuchte Temperatur des Meßobjektes aus den gemessenen Strahlungsleistungen und der bekannten Temperatur

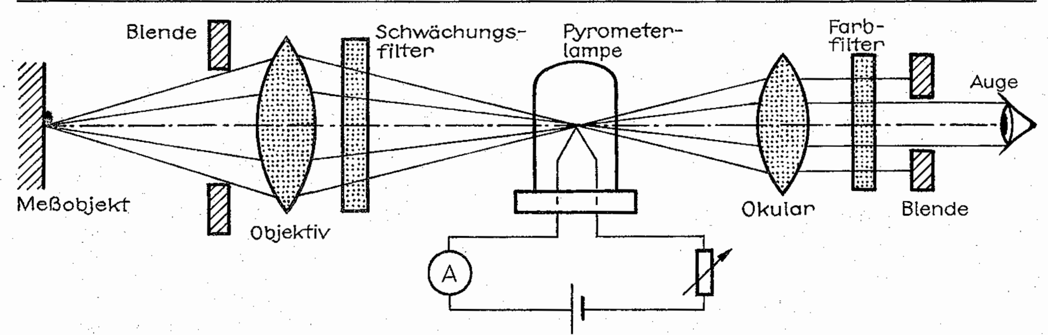


Abb. 7.3.9–8 Aufbau eines Glühfadenpyrometers

T_0 des Vergleichsstrahlers aus dem Planckschen Strahlungsgesetz (vgl. 7.5.).

Das *Quotientenpyrometer* basiert auf der Messung der spektralen Strahlungsleistung des sich auf der Temperatur T befindlichen Meßobjektes bei zwei unterschiedlichen Wellenlängen. Aus dem Verhältnis der beiden Strahlungsleistungen ergibt sich mit Hilfe des Planckschen Strahlungsgesetzes die Temperatur des Meßobjektes. Ein Vergleichsstrahler ist nicht erforderlich.

Durch den Einsatz geeigneter infrarotempfindlicher Strahlungsempfänger lassen sich mit den objektiven Pyrometern Temperaturmessungen bis herunter zu ≈ 200 K ausführen. Das ist vor allem für solche Fälle von Interesse, wo eine berührungslose Temperaturmessung erforderlich ist, denn alle anderen für den Temperaturbereich unterhalb 1 000 K bekannten Thermometer (vgl. 7.3.9.2. bis 7.3.9.6.) erfordern einen direkten thermischen Kontakt zwischen Meßobjekt und Meßfühler.

7.3.9.8. Rauschthermometer. Das Widerstandsrauschen (Nyquist-Rauschen) ist temperaturabhängig und kann demzufolge für die Temperaturmessung ausgenutzt werden. Für das mittlere Rauschspannungsquadrat gilt die einfache Beziehung $\overline{U^2} = 4kTR\,\Delta f$, wobei k die Boltzmann-Konstante, T die thermodynamische Temperatur, R die Größe des Widerstandes und Δf das Frequenzintervall ist, in dem das Rauschspannungsquadrat gemessen wird. Bei der praktischen Konstruktion von Rauschthermometern wird das mittlere Rauschspannungsquadrat $\overline{U^2}$ an dem auf der gesuchten Temperatur T befindlichen Meßwiderstand R mit dem mittleren Rauschspannungsquadrat $\overline{U_0^2}$ an einem veränderlichen Vergleichswiderstand R_0 verglichen, der sich auf einer bekannten Bezugstemperatur T_0 befindet. Dann wird R_0 so eingestellt, daß $\overline{U^2} = \overline{U_0^2}$ wird. Dann ist die gesuchte Temperatur durch die einfache Beziehung $T = T_0R_0/R$ gegeben. Rauschthermometer können unter Einsatz entsprechender elektronischer Meßmittel im Temperaturbereich $T = 1$ bis 3 000 K verwendet werden.

7.3.9.9. Magnetisches Thermometer. Im magnetischen Thermometer wird die Temperaturabhängigkeit der magnetischen Suszeptibilität paramagnetischer Salze ausgenutzt, die in vielen Fällen gut durch das Curie-Weiss-Gesetz $\chi_m = C/(T + \Delta)$ beschrieben werden kann. Die Größen C und Δ sind Materialkonstanten. Mit den magnetischen Thermometern sind Temperaturmessungen unterhalb ≈ 30 K möglich. Meist werden sie jedoch nur im Temperaturbereich unterhalb 1 K eingesetzt, da für Temperaturen oberhalb 1 K andere einfachere Thermometer zur Verfügung stehen, z. B. die Halbleiterwiderstandsthermometer (vgl. 7.3.9.5.). Als paramagnetisches Material wird vor allem Cerium-Magnesium-Nitrat $[Ce_2Mg_3(NO_3)_{12} \cdot 24\,H_2O]$ verwendet, mit dem Temperaturmessungen bis herunter zu $\approx 0{,}002$ K ausgeführt werden können. Die Messung der magnetischen Suszeptibilität erfolgt mit der Induktionsmethode (vgl. 7.4.2.).

7.3.10. Messung von Wärmemengen (Kalorimetrie)

7.3.10.1. Kalorimeter mit Messung einer Temperaturdifferenz. Die Messung einer Wärmemenge läßt sich auf die Messung einer Temperaturdifferenz zurückführen. Dabei kann es sich um die Messung einer zeitlichen oder einer räumlichen Temperaturdifferenz handeln. Die Messung einer zeitlichen Temperaturdifferenz wird im Flüssigkeits- und im Metallkalorimeter ausgenutzt. In beiden Geräten wird einem Wärmeempfänger, der zu Beginn der Messung die Temperatur T_A hat, die zu messende Wärmemenge Q zugeführt. Dadurch steigt die Temperatur des Wärmeempfängers an und wird nach Ablauf einer bestimmten Zeit, die für die Einstellung des thermischen Gleichgewichts erforderlich ist, einen konstanten Endwert T_E annehmen. Die zu messende Wärmemenge ist dann durch $Q = C_K(T_E - T_A)$ gegeben, wobei C_K die Wärmekapazität des Wärmeempfängers ist. Bei der Konstruktion der Kalorimeter ist darauf zu achten, daß eine Verfälschung der Messung durch Wärmeübertragungen zwischen dem Wärmeempfänger und der Umgebung verhindert wird. Die Wärmekapazität C_K des Wärmeempfängers wird durch Kalibrierung bestimmt, indem man dem Wärmeempfänger eine bekannte Wärmemenge zuführt (etwa durch elektrische Heizung) und die zugehörige Temperaturänderung $T_E - T_A$ mißt.

Flüssigkeitskalorimeter. Der prinzipielle Aufbau ist in Abb. 7.3.10–1 gezeigt. Als Wärmeempfänger dient die Kalorimeterflüssigkeit F (meist Wasser), die sich in dem aus dünnem Blech (vernickeltes Kupfer, Silber, Platin) hergestellten Kalorimetergefäß K befindet. Durch das Rührwerk R wird die Flüssigkeit gleichmäßig durchmischt, so daß sie überall die gleiche Temperatur hat. Zur thermischen Abschirmung gegenüber Umgebungseinflüssen dient der doppelwandige Kupferbehälter M, dessen Zwischenraum mit Wasser gefüllt ist, das mit Hilfe eines Thermostaten auf einer konstanten Temperatur gehalten wird, die nahe der Versuchstemperatur liegt. Durch einen Deckel D aus starkem Kupferblech, der in thermischem Kontakt mit dem doppelwandigen Außengefäß M steht, erfolgt die thermische Abschirmung an der Oberseite der Anordnung. Um die thermische Isolation zwischen dem Kalorimetergefäß K und dem Außengefäß M zu sichern, wird das Kalorimeter-

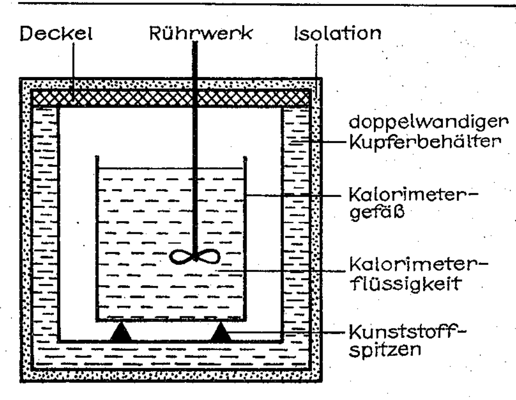


Abb. 7.3.10–1 Schema eines Flüssigkeitskalorimeters

gefäß auf einer wärmeisolierenden Unterlage (z. B. Kunststoffspitzen S) aufgestellt. Der Raum zwischen Kalorimetergefäß und Außengefäß enthält gewöhnlich Luft, kann aber bei entsprechend veränderter Konstruktion des Kalorimeters zur Verbesserung der Wärmeisolation auch evakuiert werden. Die Temperatur der Kalorimeterflüssigkeit wird mit Flüssigkeitsthermometern, Platin-Widerstandsthermometern oder Thermoelementen (vgl. 7.3.9.) gemessen.

Ein Anwendungsgebiet der Flüssigkeitskalorimeter ist die Bestimmung von Reaktionswärmen. Dazu wird in die Kalorimeterflüssigkeit ein Reaktionsgefäß gebracht, in dem die zu untersuchende Reaktion abläuft. Das Reaktionsgefäß kann vollständig geschlossen sein oder über Zuleitungen mit der Umgebung in Verbindung stehen, um die reagierenden Substanzen zu- bzw. abführen zu können. Aus der sich infolge der Reaktion ergebenden Temperaturänderung der Kalorimeterflüssigkeit läßt sich die Reaktionswärme bestimmen. Der wichtigste Typ von Kalorimetern mit Reaktionsgefäß ist das *Verbrennungskalorimeter*, in dessen geschlossenem Reaktionsgefäß (*kalorimetrische Bombe*) feste oder flüssige Stoffe nach elektrischer Zündung in überschüssigem Sauerstoff verbrannt werden. Auf diese Weise wird der spezifische Brennwert dieser Stoffe ermittelt.

Metallkalorimeter verwenden anstelle einer Kalorimeterflüssigkeit ein gut wärmeleitendes Metall (Kupfer, Silber, Aluminium) als Wärmeempfänger. Von seinem prinzipiellen Aufbau her unterscheidet sich das Metallkalorimeter nicht vom Flüssigkeitskalorimeter. Man ersetzt lediglich das Kalorimetergefäß mit der in ihm enthaltenen Kalorimeterflüssigkeit durch einen meist zylindrischen Metallblock, der in der Mitte eine Bohrung enthält, in die Versuchskörper eingeführt werden können. Die Temperaturmessung erfolgt bei den Metallkalorimetern mit einem geeigneten Thermoelement, dessen eine Lötstelle in den Metallblock eingelassen wird, während sich die andere Lötstelle wie üblich auf einer bekannten konstanten Bezugstemperatur befindet (vgl. 7.3.9.6.).

Auf der Messung einer örtlichen Temperaturdifferenz basieren das Wärmeleitungskalorimeter und das Strömungskalorimeter. Der prinzipielle Aufbau eines *Wärmeleitungskalorimeters* ist in Abb. 7.3.10–2 dargestellt. Der wesentliche Unterschied zum in Abb. 7.3.10–1 gezeigten Flüssigkeitskalorimeter besteht darin, daß hier das Kalorimetergefäß nicht wärmeisoliert zum Außengefäß aufgestellt ist, sondern daß Kalorimetergefäß und Außengefäß durch einen wärmeleitenden Stab miteinander verbunden sind. Das bedeutet, daß sich der aus dem Kalorimetergefäß und der Kalorimeterflüssigkeit (oder auch aus einem Metallblock) bestehende Wärmeempfänger auf derselben Temperatur wie das Wasser im doppelwandigen Außengefäß befindet, solange dem Wärmeempfänger keine Wärmemenge zugeführt wird. Wie beim Flüssigkeitskalorimeter wird diese Temperatur durch einen Thermostaten während der gesamten Messung konstant gehalten. Führt man jetzt dem Wärmeempfänger eine zu messende Wärmemenge Q zu, so wird sich seine Temperatur ändern, was einen Wärmestrom durch den wärmeleitenden Stab vom Wärmeempfänger zum Außengefäß zur Folge hat. Dieser Wärmestrom ist zu jedem Zeitpunkt t der Temperaturdifferenz $\Delta T = T(x_1, t) - T(x_2, t)$ zwischen den Temperaturen $T(x_1, t)$ und $T(x_2, t)$ proportional, die an zwei beliebig wählbaren benachbarten Querschnitten x_1 und x_2 des wärmeleitenden Stabes gleichzeitig gemessen werden. Diese Temperaturdifferenz $\Delta T(t)$ wird vom Beginn des Experiments zum Zeitpunkt t_A bis zum Zeitpunkt t_E ermittelt, wo die gesamte Anordnung wieder die Ausgangstemperatur angenommen hat, das heißt. $\Delta T(t) = \Delta T(t_E) = 0$ geworden ist. Dann ist die gesamte dem Wärmeempfänger zugeführte Wärmemenge Q durch den Stab geströmt. Sie ergibt sich aus der Beziehung

$$Q = K \int_{t_A}^{t_E} \Delta T(t)\, dt,$$

wobei der Proportionalitätsfaktor K eine Gerätekonstante ist, die ebenso wie die Wärmekapazität C_K beim Flüssig-

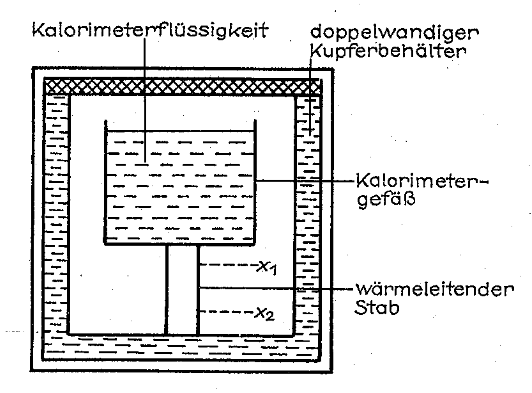


Abb. 7.3.10–2 Schema eines Wärmeleitungskalorimeters

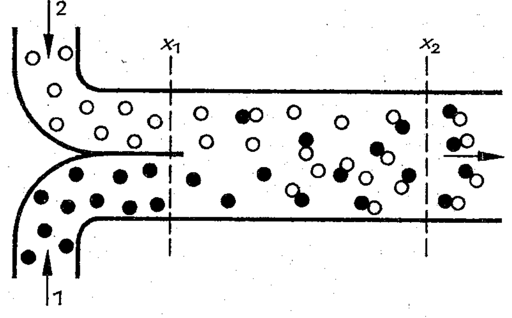


Abb. 7.3.10–3 Schema eines Strömungskalorimeters

keits- oder Metallkalorimeter durch Kalibrierung bestimmt werden muß.
Das *Strömungskalorimeter* kann zur Bestimmung der Reaktionswärme von Flüssigkeiten oder Gasen verwendet werden und ist schematisch in Abb. 7.3.10–3 dargestellt. Zwei Flüssigkeiten (oder Gase) 1 und 2, die miteinander reagieren und deren Reaktionswärme ermittelt werden soll, haben an der Meßstelle x_1, wo sie in das Reaktionsrohr einströmen, die gleiche Temperatur $T(x_1)$. Auf dem Weg von der Meßstelle x_1 zur Meßstelle x_2 reagieren sie miteinander, wobei der Abstand zwischen x_1 und x_2 so gewählt wird, daß die Reaktion bei x_2 abgeschlossen ist. Die Temperatur $T(x_2)$ bei x_2 unterscheidet sich aufgrund der auftretenden Reaktionswärme von der Temperatur $T(x_1)$ bei x_1. Strömen die Flüssigkeiten oder Gase kontinuierlich durch das Reaktionsrohr, so wird sich ein thermischer Gleichgewichtszustand einstellen. Die Reaktionswärme ist dann der Temperaturdifferenz $T(x_1) - T(x_2)$ proportional.

7.3.10.2. Phasenumwandlungskalorimeter. Die Überführung eines Stoffes aus der festen in die flüssige Phase bzw. aus der flüssigen in die gasförmige Phase erfordert eine bestimmte Umwandlungswärme (Schmelzwärme Q_s bzw. Verdampfungswärme Q_d). Während der Phasenumwandlung bleibt die Substanz auf konstanter Temperatur (Schmelztemperatur T_s bzw. Siedetemperatur T_d). Die Temperatur ändert sich erst wieder, wenn der Umwandlungsvorgang vollständig abgeschlossen ist. Dieser Effekt wird in den Phasenumwandlungskalorimetern angewendet. Die dem Kalorimeter zugeführte zu messende Wärmemenge Q überführt eine bestimmte Menge der Kalorimetersubstanz, die sich auf der entsprechenden Umwandlungstemperatur befindet, von einer Phase in eine andere. Ist m die Masse der umgewandelten Kalorimetersubstanz und Q_l die spezifische Umwandlungswärme der gewählten Substanz für den entsprechenden Phasenübergang (fest-flüssig oder flüssig-gasförmig), so ist die zu messende Wärmemenge Q durch $Q = Q_l m$ gegeben. Die Messung einer Wärmemenge wird somit auf die Messung einer Masse zurückgeführt.
Eiskalorimeter nutzen die Umwandlung von Eis in Wasser, basieren also auf einem Phasenübergang fest-flüssig. Der prinzipielle Aufbau eines Eiskalorimeters ist in Abb. 7.3.10–4 dargestellt. Es besteht aus einem Glasgefäß G, in das ein Glasrohr R eingeschmolzen ist. Der untere Teil des Glasgefäßes setzt sich mit einem Rohr fort, das in einer mit einer Strichskale versehenen Kapillaren K endet. Der obere Teil des Glasgefäßes ist mit reinem luftfreien Wasser gefüllt, der untere Teil des Gefäßes, das sich darin anschließende Rohr und die Kapillare enthalten Quecksilber. Die gesamte Anordnung ist von einem Eis-Wasser-Gemisch EW umgeben. Vor Versuchsbeginn wird in das Rohr R ein Kühlmittel gegeben, mit dessen Hilfe ein das Rohr R umgebender Eismantel E erzeugt wird. Nachdem sich Gleichgewicht eingestellt hat, befindet sich das System auf der Schmelztemperatur des Eises. Wird nun die zu messende Wärmemenge über das Rohr R zugeführt, so wird ein Teil des Eismantels schmelzen. Zur Bestimmung der geschmolzenen Eismasse wird die Tatsache ausgenutzt, daß Eis und Wasser unterschiedliche Dichte haben, d. h. beim Phasenübergang Eis-Wasser eine Volumenänderung der im Glasgefäß enthaltenen Eis-Wasser-Menge auftritt, die der Masse des geschmolzenen Eises proportional ist. Diese Volumenänderung wird aus der Änderung des Standes der Quecksilbersäule in der Kapillaren K bestimmt.
Verdampfungskalorimeter beruhen auf dem Phasenübergang flüssig-gasförmig. In ihnen wird die zu messende Wärmemenge einer auf ihrer Siedetemperatur gehaltenen Flüssigkeit zugeführt, wodurch eine bestimmte Flüssigkeitsmasse verdampft, die entweder wie beim Eiskalorimeter über eine Volumenmessung oder direkt durch Wägung bestimmt wird. Als Kalorimeterflüssigkeiten werden verflüssigte Gase (Stickstoff, Sauerstoff, Wasserstoff), flüssiger Ammoniak und organische Flüssigkeiten verwendet.

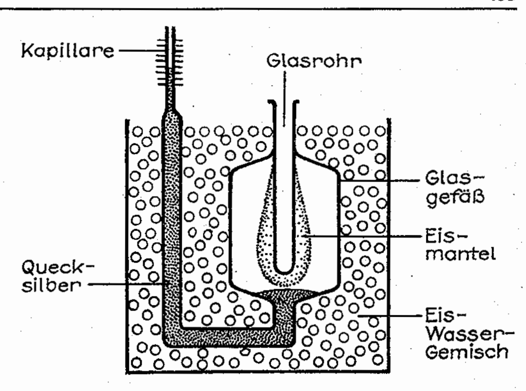


Abb. 7.3.10–4 Eiskalorimeter

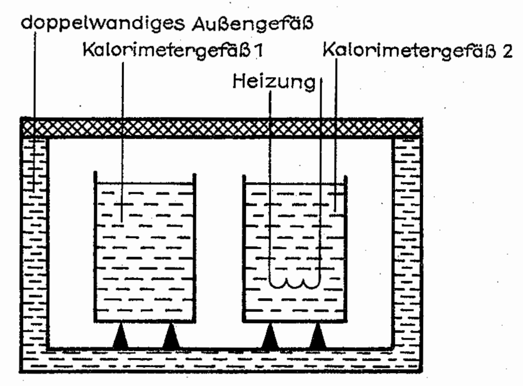


Abb. 7.3.10–5 Flüssigkeitskalorimeter in Zwillingsbauweise

7.3.10.3. Zwillingskalorimeter (Differentialkalorimeter). Zur Erhöhung der Empfindlichkeit und der Meßgenauigkeit werden Kalorimeter, die auf der Messung einer Temperaturdifferenz basieren (Flüssigkeits-, Metall- und Wärmeleitungskalorimeter, vgl. 7.3.10.1.), als Zwillingskalorimeter aufgebaut. Das Schema eines Flüssigkeitskalorimeters in Zwillingsbauweise ist in Abb. 7.3.10–5 dargestellt. Es besteht aus zwei Kalorimetergefäßen K_1 und K_2 von gleicher Beschaffenheit, die thermisch voneinander isoliert in demselben sich auf konstanter Temperatur befindlichen Außengefäß M aufgestellt sind. Dem Kalorimetergefäß K_1 wird die zu messende Wärmemenge zugeführt. Sie führt zu einer Temperaturänderung von der für beide Kalorimetergefäße gleichen Anfangstemperatur T_A auf eine Endtemperatur $T_E(K_1)$. Dem Kalorimetergefäß K_2 wird nun über die elektrische Heizung H ebenfalls Wärmeenergie zugeführt, und zwar solange, bis im Kalorimetergefäß K_2 die gleiche Endtemperatur erreicht ist wie im Kalorimetergefäß K_1, d. h. $T_E(K_2) = T_E(K_1)$ geworden ist. Dann ist die zu messende Wärmeenergie Q gleich der auf elektrischem Wege zugeführten Wärmeenergie Q_{el}, die leicht und genau gemessen

werden kann. Nach dem Jouleschen Gesetz gilt $Q_{el} = UIt$, wobei U die an der Heizung anliegende elektrische Spannung, I der durch die Heizung fließende elektrische Strom und t die Zeit ist, während welcher die Heizung betrieben werden muß, um die gewünschte Endtemperatur im Kalorimetergefäß K_2 zu erreichen. Zur richtigen Dosierung der elektrisch zugeführten Wärmeenergie ist die Temperaturdifferenz zwischen den beiden Kalorimetergefäßen zu messen. Das geschieht mit Thermoelementen, wobei die Differenz in den sich an K_1 und K_2 ergebenden Thermospannungen für die elektronische Regelung der Heizung H dient.

Metallkalorimeter in Zwillingsbauweise unterscheiden sich von dem in Abb. 7.3.10–5 schematisch dargestellten Flüssigkeitskalorimeter nur dadurch, daß die beiden Kalorimetergefäße K_1 und K_2 mit den in ihnen enthaltenen Kalorimeterflüssigkeiten durch zwei Metallblöcke gleicher Beschaffenheit ersetzt werden.

Ein Wärmeleitungskalorimeter in Zwillingsbauweise wird in Abb. 7.3.10–6 gezeigt. Hier sind die beiden gleichartigen Kalorimetergefäße K_1 und K_2 mit zwei wärmeleitenden Stäben W_1 und W_2 gleicher Beschaffenheit mit dem auf konstanter Temperatur befindlichen Außengefäß M verbunden. Die zu messende Wärmemenge Q wird dem Kalorimetergefäß K_1 zugeleitet. Sie führt während einer bestimmten Zeit von t_A bis t_E zu einem Wärmestrom vom Kalorimetergefäß K_1 über den wärmeleitenden Stab W_1 zum Außengefäß M, dessen Größe zu jedem Zeitpunkt t der Temperaturdifferenz $\Delta T(t, W_1)$ zwischen zwei benachbarten Querschnitten x_1 und x_2 des wärmeleitenden Stabes proportional ist. Beim normalen Wärmeleitungskalorimeter wird aus dieser Temperaturdifferenz die zu messende Wärmemenge Q ermittelt (vgl. 7.3.10.1.). Im Zwillingskalorimeter wird nun dem Kalorimetergefäß K_2 mit Hilfe der Heizung H ständig eine solche Wärmemenge zugeführt, daß am wärmeleitenden Stab W_2 zu jedem Zeitpunkt die gleiche Temperaturdifferenz $\Delta T(t, W_2) = \Delta T(t, W_1)$ zwischen den Querschnitten x_1 und x_2 auftritt wie am Stab W_1. Dann ist die dem Kalorimetergefäß K_2 insgesamt zugeführte Wärmemenge gleich der zu messenden Wärmemenge. Die elektrisch erzeugte Wärmemenge kann leicht und genau bestimmt werden, wenn man noch berücksichtigt, daß Heizspannung und -strom zeitlich veränderlich sein können. Es gilt

$$Q_{el} = \int_{t_A}^{t_E} U(t)I(t)\,dt.$$

Zur Regelung der Heizung dient die Temperaturdifferenz $\Delta T(t) = \Delta T(t, W_1) - \Delta T(t, W_2)$, die mit Hilfe von Thermoelementen gemessen wird.

7.3.10.4. Scanning-Kalorimeter. Die bisher besprochenen Kalorimeter (vgl. 7.3.10.1. bis 7.3.10.3.) zeichnen sich dadurch aus, daß sie bei einer festen Arbeitstemperatur betrieben werden. Sie ist im Falle der auf Messung einer Temperaturdifferenz basierenden Kalorimeter durch die Temperatur des Außengefäßes gegeben, bei den Phasenumwandlungskalorimetern wird sie durch die Schmelz- bzw. Siedetemperatur der Kalorimetersubstanz bestimmt. Insbesondere bei der Bestimmung thermischer Stoffgrößen (vgl. 7.3.11.) ist es häufig von Interesse, deren Temperaturabhängigkeit zu ermitteln. Außerdem treten manche Prozesse (z. B. Phasenumwandlungen) nur bei bestimmten Temperaturen auf, so daß es für die Untersuchung der damit verbundenen thermischen Effekte erforderlich ist, über ein Kalorimeter mit entsprechender Arbeitstemperatur zu verfügen. Für derartige Messungen sind die Scanning-Kalorimeter entwickelt worden. Man unterscheidet drei Arten des Scanning-Betriebs:

1. Umgebungs-Scanning-Betrieb. Diese Betriebsart wird bei Wärmeleitungskalorimetern angewendet. Dabei wird die Temperatur des Außengefäßes M linear mit der Zeit erhöht, $T_M = T_{M0} + at$ (a = Heizrate), was durch elektronische Regelung einer entsprechend angebrachten elektrischen Heizung geschieht. Das Meßsystem, bestehend aus Kalorimetergefäß und Kalorimeterflüssigkeit oder aus einem Metallblock, folgt dann aufgrund des entstehenden Wärmestroms durch den wärmeleitenden Stab W mit einer gewissen zeitlichen Verzögerung dieser Temperaturerhöhung im Außengefäß (vgl. Abb. 7.3.10–2).

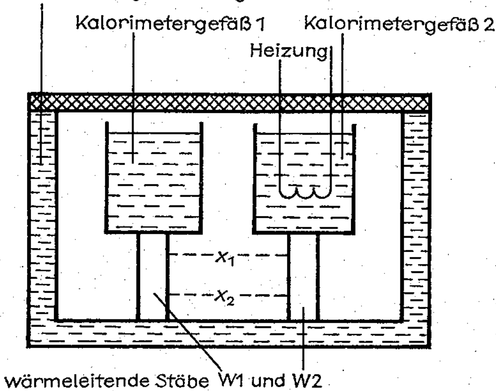


Abb. 7.3.10–6 Wärmeleitungskalorimeter in Zwillingsbauweise

2. Adiabatischer Scanning-Betrieb. Bei dieser Betriebsart werden sowohl das Außengefäß als auch die Meßzelle zeitlinear aufgeheizt, und zwar über zwei separat geregelte Heizungen. Die Aufheizung erfolgt dabei so, daß die Temperatur T_M des Außengefäßes stets gleich der Temperatur T_K der Meßzelle ist, $T_M(t) = T_K(t)$.

3. Isoperiboler Scanning-Betrieb. Bei dieser Betriebsart wird die Temperatur des Außengefäßes konstant gehalten, während die Meßzelle zeitlinear aufgeheizt wird, $T_K(t) = T_{K0} + at$.

Scanning-Kalorimeter werden in der Regel als Zwillingskalorimeter aufgebaut (vgl. 7.3.10.3.). Der überstrichene Temperaturbereich erstreckt sich von ≈80 K (Kühlung des Außengefäßes mit flüssigem Stickstoff) bis zu ≈2700 K. Bei hohen Temperaturen ist das Außengefäß ein Ofen, der keine Flüssigkeit mehr enthält. Als Meßzellen, die das Probenmaterial aufnehmen, dienen Metallblöcke oder Metallbehälter. Die obere Betriebstemperatur des Scanning-Kalorimeters wird durch den Schmelzpunkt des Metalls bestimmt, das für die Herstellung der Meßzellen verwendet wird.

7.3.11. Messung thermischer Stoffgrößen

7.3.11.1. Spezifische Wärmekapazität. Die spezifische Wärmekapazität c eines Stoffes ist die Wärmemenge, die benötigt wird, um 1 Kilogramm dieses Stoffes von einer

Temperatur T auf eine um 1 K höhere Temperatur zu erwärmen. Die spezifische Wärmekapazität ist eine Funktion der Temperatur. Außerdem muß man zwischen der spezifischen Wärmekapazität c_p bei konstantem Druck und der spezifischen Wärmekapazität c_v bei konstantem Volumen des Stoffes unterscheiden. In Festkörpern und Flüssigkeiten unterscheiden sich c_p und c_v nur wenig voneinander. Es gilt die Beziehung $c_p = c_v + \beta^2 T/(\varrho\chi)$, wobei β der kubische Ausdehnungskoeffizient des Stoffes, ϱ seine Dichte und χ seine Kompressibilität sind. In Gasen ist der Unterschied zwischen c_p und c_v größer und wird durch das Verhältnis $\gamma = c_p/c_v$ gekennzeichnet. Für ideale Gase gilt im Falle eines einatomigen Gases (z. B. Edelgase) $\gamma = 1{,}66$, im Falle von aus zwei Atomen bestehenden Gasmolekülen (z. B. H_2, N_2, O_2) $\gamma = 1{,}40$ und für aus mehr als zwei Atomen bestehenden Gasmolekülen $\gamma = 1{,}33$. Die Einheit der spezifischen Wärmekapazität ist $1\,\mathrm{J\,kg^{-1}\,K^{-1}}$. Häufig wird anstelle der auf die Masseneinheit (1 kg) bezogenen spezifischen Wärmekapazität die auf die Einheit der Stoffmenge (1 mol) bezogene Wärmekapazität verwendet. Sie wird als molare Wärmekapazität bezeichnet und durch große Buchstaben C_p und C_v gekennzeichnet. Ihre Einheit ist $1\,\mathrm{J\,mol^{-1}\,K^{-1}}$.

Die Messung der spezifischen Wärmekapazität erfolgt mit Hilfe von Kalorimetern. Werden keine speziellen Maßnahmen zur Konstanthaltung des Volumens des untersuchten Stoffes getroffen, so wird stets die spezifische Wärmekapazität c_p bei konstantem Druck gemessen. Die Bestimmung der spezifischen Wärmekapazität bei einer vorgegebenen Temperatur erfolgt nach der Mischungsmethode oder in einem Flüssigkeits- oder Metallkalorimeter, dessen Kalorimetergefäß mit einer zusätzlichen elektrischen Heizung versehen ist. Bei der Mischungsmethode wird der zu untersuchende Körper zur Ermittlung seiner Masse m zunächst gewogen und dann in einem Vorwärmer auf eine Anfangstemperatur T gebracht. Dann wird der Körper in das Kalorimeter gebracht. Handelt es sich dabei um ein Flüssigkeits- oder Metallkalorimeter (vgl. 7.3.10.1.), so wird die Temperatur des Wärmeempfängers von der Anfangstemperatur T_A auf eine Endtemperatur T_E ansteigen, wobei sich gleichzeitig die Temperatur des zu untersuchenden Körpers von T auf T_E verringert. Die vom Körper abgegebene Wärmemenge ist dann gleich der vom Wärmeempfänger aufgenommenen Wärmemenge, das heißt, es gilt $c_p m(T - T_E) = C_K(T_E - T_A)$, wobei C_K die Wärmekapazität des Wärmeempfängers ist. Hieraus läßt sich c_p berechnen. Wird ein Wärmeleitungskalorimeter (vgl. 7.3.10.1.) oder ein Phasenumwandlungskalorimeter (vgl. 7.3.10.2.) verwendet, so kühlt sich der zu untersuchende Körper auf die konstante Betriebstemperatur T_i des Kalorimeters ab, so daß die von ihm an das Kalorimeter abgegebene Wärmemenge gleich $c_p m(T - T_i)$ ist. Hierbei ist T_i entweder die konstant gehaltene Temperatur des Außengefäßes (Wärmeleitungskalorimeter) oder die Phasenumwandlungstemperatur (Phasenumwandlungskalorimeter).

Bei der zweiten Methode mit zusätzlicher elektrischer Heizung wird der zu untersuchende Körper der Masse m vor der Messung in das Kalorimeter gebracht, so daß sich Kalorimeter und Körper auf der gleichen Anfangstemperatur T_A befinden. Dann wird dem Kalorimeter mit Hilfe der elektrischen Heizung die Wärmeenergie $Q = UIt$ zugeführt (U = Heizspannung, I = Heizstrom, t = Heizdauer). Dadurch steigt die Temperatur des Kalorimeters von T_A auf T_E an. Da sowohl der Wärmeempfänger des Kalorimeters mit der Wärmekapazität C_K als auch der im Kalorimetergefäß befindliche zu untersuchende Körper mit der Wärmekapazität $c_p m$ aufgeheizt werden, gilt die Beziehung $UIT = (C_K + c_p m)(T_E - T_A)$, woraus c_p zu berechnen ist.

Die beiden beschriebenen Methoden lassen sich in der dargestellten Weise für die Bestimmung der spezifischen Wärmekapazität von Festkörpern und Flüssigkeiten einsetzen. Zur Ermittlung der spezifischen Wärmekapazität c_p bei konstantem Druck von Gasen läßt sich die Mischungsmethode einsetzen. Dazu verwendet man ein Flüssigkeitskalorimeter, in dessen Flüssigkeit ein Schlangenrohr angeordnet ist, durch welches das Gas hindurchströmen kann. Das Gas wird vor dem Eintritt in das Kalorimeter auf eine Temperatur T_1 erhitzt; dann leitet man es in starkem kontinuierlichem Strom durch das Schlangenrohr des Kalorimeters, wobei es sich auf die Austrittstemperatur T_2 abkühlt. Die vom Gas an das Kalorimeter abgegebene und in üblicher Weise bestimmbare Wärmemenge (vgl. 7.3.10.1.) ist dann gleich $c_p m(T_1 - T_2)$, wobei die Masse m des während einer bestimmten Zeit durch das Schlangenrohr hindurchgeströmten Gases mit einem Gasometer bestimmt wird. Zur Ermittlung der spezifischen Wärmekapazität c_v bei konstantem Volumen wird eine bestimmte Gasmenge m in einem Gefäß eingeschlossen. Dann führt man dem Gas mit Hilfe einer elektrischen Heizung (meist nur ein kurzer Stromstoß von 0,01 bis 0,1 s Dauer) eine bekannte Wärmemenge $Q = UIt$ zu und mißt direkt den daraus resultierenden Temperaturanstieg des Gases. Sind T_1 und T_2 die Anfangs- bzw. Endtemperatur des Gases, so gilt $UIt = c_v m(T_2 - T_1)$, wobei m die Masse des Gases ist. Die Messung der Temperaturabhängigkeit der spezifischen Wärmekapazität erfolgt mit Scanning-Kalorimetern (vgl. 7.3.10.4.). In Kalorimetern mit adiabatischem und isoperibolem Scanning-Betrieb werden die beiden Meßzellen K_1 und K_2 des Scanning-Kalorimeters durch zwei voneinander unabhängige elektronisch geregelte elektrische Heizungen so aufgeheizt, daß ihre Temperaturen T_1 und T_2 zu jedem Zeitpunkt t gleich groß sind und daß die Temperatur linear mit der Zeit ansteigt, $T_1(t) = T_2(t) = T(t) = T_0 + at$. Hier ist T_0 die Anfangstemperatur der Messung und die konstante Größe a wird als Heizrate bezeichnet. Der zu untersuchende Körper mit der Masse m, dessen spezifische Wärmekapazität c_p ermittelt werden soll, wird nun in die Meßzelle K_1 gebracht, während die Meßzelle K_2 leer bleibt (mitunter bringt man in die Meßzelle K_2 auch einen Vergleichskörper, dessen spezifische Wärmekapazität genau bekannt ist). Die für die Aufheizung der beiden Meßzellen erforderlichen Wärmeströme bzw. Heizleistungen $\mathrm{d}Q_1/\mathrm{d}t = U_1(t)I_1(t)$ und $\mathrm{d}Q_2/\mathrm{d}t = U_2(t)I_2(t)$ werden sich dann voneinander unterscheiden, da im Falle der Meßzelle K_1 neben der eigentlichen Meßzelle mit ihrer Wärmekapazität C_K auch noch der zu untersuchende Körper mit der Wärmekapazität $c_p m$ aufzuheizen ist. Aus der allgemeinen Beziehung $\mathrm{d}Q = c_p m\,\mathrm{d}T$ folgen bei zeitlinearem Temperaturanstieg mit der Heizrate a die Relationen $\mathrm{d}Q_1/\mathrm{d}t = (C_K + c_p m)a$ und $\mathrm{d}Q_2/\mathrm{d}t = C_K a$. Hieraus ergibt sich für die spezifische Wärmekapazität $c_p(t) = [U_1(t)I_1(t) - U_2(t)I_2(t)]/(ma)$.

Die für die Ermittlung der spezifischen Wärmekapazität benötigte Meßgröße ist folglich die Differenz der Heizleistungen. In den ohnehin mit komplizierten elektronischen Regelsystemen ausgestatteten Scanning-Kalorimetern wird sie zu jedem Zeitpunkt t auf elektronischem Wege berechnet. Die Umrechnung von $c_p(t)$ auf die Temperaturabhängigkeit der spezifischen Wärmekapazität $c_p(T)$ erfolgt unter Berücksichtigung der linearen Relation $T(t) = T_0 + at$. In einem Kalorimeter mit Umgebungs-Scanning-Betrieb (vgl. 7.3.10.4.) wird ebenso wie im vorhergehenden Fall der zu

untersuchende Körper in die Meßzelle K_1 gebracht, während die Meßzelle K_2 leer bleibt. Wird nun die Temperatur T_M des Außengefäßes zeitlinear erhöht, $T_M(t) = T_{M0} + at$, so werden mit einer gewissen zeitlichen Verzögerung auch die Temperaturen T_1 und T_2 der beiden Meßzellen ansteigen. Wegen der unterschiedlichen Gesamtwärmekapazität der beiden Meßzellen ($C_K + c_p m$ in K_1 und C_K in K_2) wird jedoch die Temperatur T_2 schneller wachsen als die Temperatur T_1. Die Temperaturdifferenz $T_2 - T_1$ ist damit ein Maß für c_p. Zur Bestimmung von c_p sind die Zeitabhängigkeit von $T_2 - T_1$, der zeitlichen Ableitung dieser Größe $\mathrm{d}(T_2 - T_1)/\mathrm{d}t$ und von T_1 zu messen. Die Formel für die Berechnung von c_p hängt von der spezifischen Konstruktion des Scanning-Kalorimeters ab. Im einfachsten Falle gilt

$$c_p = \frac{C_K}{m} \frac{C_K \,\mathrm{d}(T_2 - T_1)/\mathrm{d}t + b(T_2 - T_1)}{b(T_M - T_1) - C_K \,\mathrm{d}(T_2 - T_1)/\mathrm{d}t - b(T_2 - T_1)},$$

wobei die Konstante b die Wärmeleitungseigenschaften der wärmeleitenden Stäbe charakterisiert und durch Kalibrierung zu bestimmen ist.

7.3.11.2. Thermische Ausdehnung. Festkörper und Flüssigkeiten zeigen eine Änderung ihres Volumens mit der Temperatur, die durch den kubischen Ausdehnungskoeffizienten β beschrieben wird. Er ist über die Beziehung $V(T) = V(T_0)[1 + \beta(T - T_0)]$ definiert und hat die Dimension K^{-1}. Als Bezugstemperatur T_0 wird meist der Eispunkt des Wassers, $T_0 = 273{,}15$ K, gewählt. Der *kubische Ausdehnungskoeffizient* β ist i. allg. eine temperaturabhängige Größe. Zur Bestimmung des kubischen Ausdehnungskoeffizienten und seiner Temperaturabhängigkeit ist das Volumen des untersuchten Körpers bei verschiedenen Temperaturen zu messen. Da die auftretenden Volumenänderungen klein sind, ist ein hinreichend empfindliches und genaues Volumenmeßverfahren (vgl. 7.3.1.) zu wählen.

Feste Körper können eine anisotrope thermische Ausdehnung zeigen, d. h., zur vollständigen Charakterisierung des Ausdehnungsverhaltens von Festkörpern ist die lineare thermische Ausdehnung in drei aufeinander senkrecht stehenden Richtungen zu messen. Analog zum kubischen Ausdehnungskoeffizienten ist der *lineare thermische Ausdehnungskoeffizient* α_l über die Beziehung $l(T) = l(T_0)[1 + \alpha_l(T - T_0)]$ definiert, wobei l die lineare geometrische Abmessung des Körpers in der betrachteten Richtung ist. Sind α_{l1}, α_{l2} und α_{l3} die linearen thermischen Ausdehnungskoeffizienten in drei aufeinander senkrecht stehenden Richtungen, so gilt in erster Näherung $\beta = \alpha_{l1} + \alpha_{l2} + \alpha_{l3}$. In kubischen Kristallen und homogenen isotropen Stoffen ist $\alpha_{l1} = \alpha_{l2} = \alpha_{l3}$. Zur Messung des linearen thermischen Ausdehnungskoeffizienten werden meist stabförmige Probenkörper des zu untersuchenden Materials hergestellt, deren Länge als Funktion der Temperatur ermittelt wird. Zur Längenmessung werden Dilatometer oder Interferenzverfahren (vgl. 7.3.1.) eingesetzt. In kristallinen Festkörpern können die linearen thermischen Ausdehnungskoeffizienten auch durch Röntgen-Beugungsmethoden (vgl. 7.8.) über die Messung der Temperaturabhängigkeit der Gitterparameter ermittelt werden.

7.3.11.3. Wärmeleitung. Betrachtet man einen länglichen Körper von sonst beliebiger geometrischer Form, dessen unteres Ende durch eine Kühlvorrichtung K auf einer Temperatur T_2 gehalten und dessen oberes Ende durch eine Heizung H auf eine Temperatur $T_1 > T_2$ gebracht wird (Abb. 7.3.11–1a), so tritt in x-Richtung von H nach K ein Wärmestrom $\Phi = \mathrm{d}Q/\mathrm{d}t$ durch den Körper auf. Dieser Wärmestrom ist an jeder beliebig herausgegriffenen Stelle x_0 des Körpers seiner Querschnittsfläche $A(x_0)$ an dieser Stelle und dem dort herrschenden Temperaturgradienten proportional,

$$\Phi(x_0) = -\lambda A(x_0)\,\mathrm{d}T/\mathrm{d}x|_{x_0}.$$

Die Proportionalitätskonstante λ ist eine materialspezifische Größe und wird als *Wärmeleitfähigkeit* des Körpers bezeichnet. Die Einheit der Wärmeleitfähigkeit ist $1\,\mathrm{W\,m^{-1}K^{-1}}$.

Zur Messung der Wärmeleitfähigkeit von Festkörpern werden Versuchskörper einfacher geometrischer Form verwendet (Stäbe, Platten, Hohlzylinder oder Hohlkugeln aus dem zu untersuchenden Material). Außerdem wird unter stationären Bedingungen gemessen, d. h., die Temperaturen T_1 und T_2 und damit auch der Wärmestrom Φ sind zeitlich konstant. Unter diesen Bedingungen ergeben sich besonders einfache Beziehungen für die Ermittlung der Wärmeleitfähigkeit aus den Temperaturen T_1 und T_2 und dem Wärmestrom Φ. Im Falle eines stabförmigen Versuchskörpers der Länge l mit dem Querschnitt A (Abb. 7.3.11–1b) gilt $\lambda = \Phi l/[A(T_1 - T_2)]$. Für einen plattenförmigen Versuchskörper gilt dieselbe Beziehung, wenn man für l die Dicke der Platte einsetzt. Bei einem hohlzylindrischen Versuchskörper mit einem Innendurchmesser d_1, einem Außendurchmesser d_2 und einer Länge L (Abb. 7.3.11–1c), der innen gekühlt und außen aufgeheizt wird, erhält man

$$\lambda = \frac{\Phi}{2\pi L} \frac{\ln(d_2/d_1)}{T_1 - T_2}.$$

Wird schließlich ein hohlkugelförmiger Versuchskörper mit einem Innenradius r_1 und einem Außenradius r_2 mit Behei-

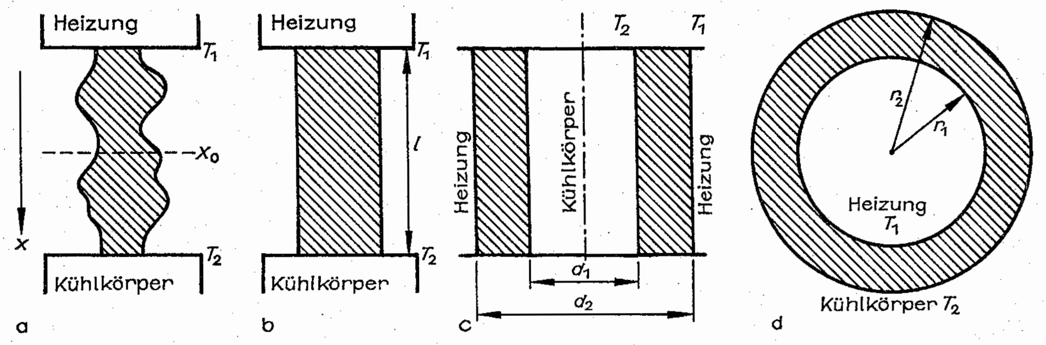


Abb. 7.3.11–1 Zur Bestimmung der Wärmeleitfähigkeit fester Körper

zung der Innenfläche und Kühlung der Außenfläche der Kugel verwendet (Abb. 7.3.11–1d), so ergibt sich

$$\lambda = \frac{r_2 - r_1}{2\pi r_1 r_2} \cdot \frac{\Phi}{T_1 - T_2}.$$

Die Temperaturen T_1 und T_2 werden mit in den Versuchskörper eingelassenen Thermoelementen (vgl. 7.3.9.6.) gemessen. Werden elektrische Heizungen verwendet, so ergibt sich der Wärmestrom Φ entsprechend dem Jouleschen Gesetz aus Heizspannung U und Heizstrom I zu $\Phi = UI$. Das Kühlsystem zur Einstellung der Temperatur T_2 ist häufig eine Flüssigkeitskühlung, die von einem Thermostaten gespeist wird.

Die für die Bestimmung der Wärmeleitfähigkeit von Flüssigkeiten am häufigsten verwendeten Methoden sind das Kapillar- oder Heizdrahtverfahren und das Lamellenverfahren.

Heizdrahtverfahren. Hierbei wird die zu untersuchende Flüssigkeit in eine Metallkapillare gefüllt, in deren Achse genau zentriert ein Heizdraht aus Platin oder Nickel ausgespannt ist. Der Heizdraht wird durch elektrische Heizung auf die Temperatur T_1 gebracht, die Kapillarenwand durch äußere Kühlung auf der konstanten Temperatur T_2 gehalten. Von der Geometrie der Anordnung her gesehen entspricht dieses System genau dem in Abb. 7.3.11–1c dargestellten Fall des hohlzylindrischen Versuchskörpers. Die Wärmeleitfähigkeit der Flüssigkeit läßt sich mit der für diesen Fall angegebenen Formel berechnen, wenn man den Innendurchmesser d_1 des Hohlzylinders durch den Außendurchmesser des Heizdrahtes und den Außendurchmesser d_2 des Hohlzylinders durch den Innendurchmesser der Kapillaren ersetzt.

Lamellenverfahren. Das Meßsystem besteht aus einer kreisförmigen Heizplatte H, die sich im Inneren eines dosenförmigen Kühlkörpers K befindet (Abb. 7.3.11–2). Die Halterung der Heizplatte erfolgt mit Hilfe eines schlecht wärmeleitenden Doppelrohres C, durch das die Heizleitungen und Thermoelementzuführungen zur Messung von T_1 laufen. Der Zwischenraum zwischen Heizplatte und Kühlkörper wird vollständig mit der zu untersuchenden Flüssigkeit F gefüllt. Bezüglich seiner Geometrie ähnelt dieses System dem in der Abb. 7.3.11–1b dargestellten Fall eines stab- oder plattenförmigen Versuchskörpers, so daß die Wärmeleitfähigkeit der Flüssigkeit aus der entsprechenden Beziehung berechnet werden kann. Dabei ist jedoch zu beachten, daß hier die Wärmeableitung von der Heizplatte zum Kühlsystem über zwei Flüssigkeitsschichten erfolgt.

Für die Messung der Wärmeleitfähigkeit von Gasen sind die beiden für Flüssigkeiten beschriebenen Verfahren einsetzbar. In der Praxis wird jedoch meist nur das Heizdrahtverfahren verwendet.

7.3.12. Feuchtemessung

7.3.12.1. Feuchte fester und flüssiger Stoffe. Unter der Feuchte eines Stoffes versteht man seinen Wassergehalt. Die gebräuchlichsten Kenngrößen zur Angabe der Feuchte in festen Stoffen sind der Feuchteanteil und der Feuchtesatz. Der *Feuchteanteil* f ist definiert als das Verhältnis der in einem Stoff erhaltenen Wassermasse m_w zur Gesamtmasse m des feuchten Stoffes, $f = m_w/m$. Der *Feuchtesatz* u ist das Verhältnis der im Stoff enthaltenen Wassermasse m_w zur Masse $m_0 = m - m_w$ des wasserfreien Stoffes, $u = m_w/m_0$. Beide Größen werden in Prozent angegeben.

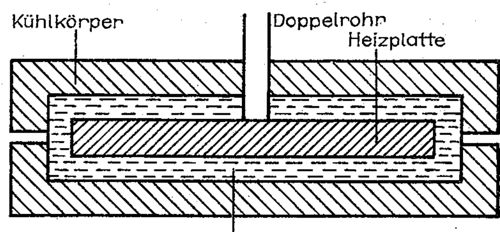


Abb. 7.3.11–2 Experimentelle Anordnung beim Lamellenverfahren

In der Feuchtemessung von festen und flüssigen Stoffen unterscheidet man zwischen direkten und indirekten Verfahren.

Direkte Feuchtemeßverfahren. Bei ihnen wird die Masse m_w des im Stoff enthaltenen Wassers direkt bestimmt. Die am häufigsten verwendete direkte Methode zur Bestimmung der Feuchte fester Stoffe ist das *Wäge-Trocknungs-Verfahren*. Dabei wird von einer Probe des zu untersuchenden Stoffes zunächst die Masse m bestimmt. Dann wird sie solange getrocknet, bis sich ihre Masse nicht mehr ändert, d. h., gleich $m_0 = m - m_w$ geworden ist. Die Trocknungsbedingungen sind stoffspezifisch zu wählen und sind in entsprechenden Vorschriften festgelegt. Die direkte Feuchtemessung von Flüssigkeiten erfolgt entweder durch Destillation mit anschließender Wägung der entstandenen Wassermenge oder über chemische Reaktionen. Das *Calciumcarbidverfahren* nutzt die Reaktion

$$CaC_2 + 2\,H_2O \rightarrow Ca(OH)_2 + C_2H_2,$$

das *Calciumhydridverfahren* die Reaktion

$$CaH_2 + 2\,H_2O \rightarrow Ca(OH)_2 + 2\,H_2$$

und das *Karl-Fischer-Verfahren* die Reaktion

$$J_2 + SO_2 + 2\,H_2O \rightarrow H_2SO_4 + 2\,HJ$$

aus. Die Menge des entstehenden Ethins C_2H_2 oder Wasserstoffes H_2 bzw. die Menge der für die Umsetzung des in der Probe enthaltenen Wassers erforderlichen Karl-Fischer-Lösung ($J_2 + SO_2$) dienen als Maß für die umgesetzte Wassermasse.

Indirekte Feuchtemeßverfahren basieren auf der Tatsache, daß die elektrische Leitfähigkeit und die dielektrischen Eigenschaften (Dielektrizitätskonstante, Verlustwinkel) fester und flüssiger Stoffe eine deutlich ausgeprägte Abhängigkeit von der Feuchte zeigen. Sowohl die elektrische Leitfähigkeit als auch die dielektrischen Parameter können bestimmt werden, indem man wenigstens zwei Elektroden in den zu untersuchenden Stoff hineinbringt und die elektrische Leitfähigkeit (oder den elektrischen Widerstand) des Stoffes zwischen diesen Elektroden bzw. die Kapazität dieser Elektrodenanordnung mißt. Bei Flüssigkeiten, Schüttgütern (z. B. Getreide, Düngemittel) und Stoffen mit lockerer, fasriger Struktur (z. B. Rohbaumwolle, Glaswolle, Watte) werden meist Zylinder- oder Plattenkondensatoren als Meßwertaufnehmer verwendet, die in das Meßgut hineingebracht werden. Im Falle kompakter Feststoffe werden die Elektroden auf die Oberfläche des Stoffes aufgesetzt. Der Einsatz dieser indirekten Feuchtemeßverfahren ist nur dann möglich, wenn für den zu untersuchenden Stoff die Abhängigkeit seiner elektrischen und dielektrischen Eigenschaften von der Feuchte bekannt ist. Die Messung des

elektrischen Widerstandes oder der Kapazität erfolgt nach einem der in 7.4.1.3. beschriebenen Verfahren.

Eine Zwischenstellung zwischen den auf einer Wägung basierenden direkten Feuchtemeßverfahren und den auf einer elektrischen Messung basierenden indirekten Feuchtemeßverfahren nehmen die spektroskopischen Feuchtemeßverfahren ein. Die Ausnutzung der magnetischen Kernresonanz beruht auf der Resonanzabsorption von Hochfrequenzenergie durch die Protonen in einem starken äußeren Magnetfeld. Die integrale Intensität der mit einem Kernresonanzspektrometer aufgenommenen Resonanzlinie ist ein Maß für die Anzahl der in der untersuchten Probe enthaltenen Protonen.

Optische Feuchtemeßverfahren beruhen auf dem Vorhandensein intensiver Absorptionsbanden des Wassers im Wellenlängenbereich von 0,8 bis 3 µm. Mit Hilfe eines geeigneten optischen Spektrometers (vgl. 7.5.) wird die Absorptionskurve einer dieser Banden gemessen. Die integrale Absorption dieser Bande ist ein Maß für die Anzahl der Wassermoleküle im untersuchten Stoff.

7.3.12.2. Gasfeuchtemessung (Hygrometrie). Zur quantitativen Angabe der Gasfeuchte werden unterschiedliche Größen verwendet. Die *absolute Gasfeuchte* d_v ist die in einer Volumeneinheit des feuchten Gases enthaltene Wasserdampfmasse. Sie wird in $kg\,m^{-3}$ gemessen. Der *Dampfdruck* e ist der Partialdruck des Wasserdampfes im feuchten Gas und wird in Pa gemessen. Der Dampfdruck e kann zwischen $e = 0$ (trockenes Gas) und dem für eine gegebene Temperatur charakteristischen Sättigungsdampfdruck e_w variieren. Als *Mischungsverhältnis eines feuchten Gases* r bezeichnet man das Verhältnis der Masse m_v des im Gas enthaltenen Wasserdampfes zur Masse m_a des trockenen Gases, $r = m_v/m_a$. Die *spezifische Feuchte* q ist das Verhältnis der Masse m_v des im feuchten Gas enthaltenen Wasserdampfes zur Gesamtmasse m des feuchten Gases, $q = m_v/m$. Die Größen r und q entsprechen bezüglich ihrer Definition dem Feuchtesatz bzw. Feuchteanteil, die für die Kennzeichnung der Feuchte fester und flüssiger Stoffe verwendet werden (vgl. 7.3.12.1.). Der *molare Wasserdampfanteil* x_v ist das Verhältnis des Stoffmenge des Wasserdampfes zur Gesamtstoffmenge des feuchten Gases (beide Stoffmengen in mol gemessen). Als *Taupunkttemperatur* T_d eines feuchten Gases wird die Temperatur bezeichnet, bei der das Gas bei gleichem Druck und gleichem Mischungsverhältnis mit Wasserdampf gesättigt ist. Die sehr häufig verwendete *relative Feuchte U* ist das Verhältnis des molaren Wasserdampfanteils des feuchten Gases zum molaren Wasserdampfanteil im Fall der Sättigung bei den gleichen Werten für den Druck und die Temperatur. Unter Ausnutzung der Zustandsgleichung der Gase lassen sich alle hygrometrischen Kenngrößen ineinander umrechnen. Anwendungszweck und Meßverfahren bedingen die bevorzugte Verwendung der einen oder anderen Kenngröße.

Für die Gasfeuchtemessung ist eine Vielzahl von Verfahren entwickelt worden, die auf den unterschiedlichsten physikalischen Wirkprinzipien basieren.

Psychrometer beruhen auf dem Zusammenhang zwischen der Gasfeuchte und der Differenz in den Temperaturen T_t und T_f, die ein trockenes bzw. feuchtes Thermometer in diesem Gas anzeigen. Aus dieser sogenannten psychrometrischen Differenz $T_t - T_f$ ergibt sich der Dampfdruck e des feuchten Gases mit Hilfe der Formel $e = e_w(T_f) - Ap(T_t - T_f)$, wobei $e_w(T_f)$ der Sättigungsdampfdruck des Wasserdampfes bei der Temperatur T_f, p der Gasdruck und A die Psychrometerkonstante ist, die experimentell durch Kalibrierung zu bestimmen ist. Die Psychrometerkonstante hängt von der Gasart ab. Die Temperaturmessung erfolgt mit Flüssigkeitsthermometern, Thermoelementen, Widerstandsthermometern oder mit Halbleiterbauelementen (vgl. 7.3.9.). Zur Bestimmung von T_t wird ein Thermometer in das Gas gebracht, die Messung von T_f erfolgt mit einem zweiten Thermometer, dessen Meßfühler mit Wasser bedeckt sein muß. Häufig umgibt man dazu den Meßfühler mit einem feuchten Strumpf aus dünnem Baumwollgewebe.

Taupunkthygrometer dienen der Ermittlung der Gasfeuchte über die Messung der Taupunkttemperatur. Dazu wird ein in das Gas gebrachter Körper solange abgekühlt, bis an seiner Oberfläche eine Kondensation des Wasserdampfes eintritt. Als Kondensationsflächen (Tauspiegel) dienen polierte Flächen aus Silber, Nickel, Edelstahl oder anderen korrosionsbeständigen Materialien, die teilweise noch mit einer Antikorrosionsschicht aus Rhodium oder Gold bezogen werden. Die Kühlung erfolgt in der Regel mit Hilfe thermoelektrischer Kühlelemente unter Ausnutzung des Peltier-Effektes (vgl. 2.6.11.5., 8.1.7.2.), die Temperaturmessung mit Thermoelementen oder Widerstandsthermometern (vgl. 7.3.9.). Der Nachweis der beginnenden Wasserdampfkondensation auf dem Tauspiegel kann in unterschiedlicher Weise erfolgen. Optische Verfahren nutzen die sprunghafte Abnahme des Reflexionsvermögens der Oberfläche für sichtbares Licht aus, die bei der Kondensation von Wasserdampf auftritt. Elektrische Methoden beruhen auf der Tatsache, daß sich der Oberflächenwiderstand der Kondensationsfläche bei der Absorption von Wasserdampf ändert. Die Absorption von α- oder β-Strahlung durch die Kondensatschicht ist Grundlage der radiometrischen Meßverfahren. Dazu wird der Tauspiegel mit einer sehr dünnen Schicht der als Strahlungsquelle verwendeten Isotope (meist das Isotop ^{210}Po für α-Strahlung und ^{63}Ni für β-Strahlung) bedampft.

Lithiumchloridtaupunktfühler. In diesem Gerät dient eine gesättigte wäßrige Lösung von Lithiumchlorid (LiCl) als feuchteempfindliches Element. Hat das Gas die Temperatur T und ist der Dampfdruck $e_{LiCl}(T)$ der Salzlösung bei dieser Temperatur kleiner als der Dampfdruck e des feuchten Gases, so kann durch Erhöhung der Temperatur der Salzlösung ein Gleichgewichtszustand zwischen dem Dampfdruck der Salzlösung und dem Dampfdruck des Gases hergestellt werden. Für die Gleichgewichtstemperatur T_g gilt dann $e = e_{LiCl}(T_g)$, die relative Feuchte des Gases ist $U = e_{LiCl}(T_g)/\acute{e}_w(T)$, und die Taupunkttemperatur läßt sich aus der Bedingung $e_{LiCl}(T_g) = e_w(T_d)$ bestimmen.

Der Lithiumchloridtaupunktfühler besteht aus einem Trägerrohr aus Glas oder aus einem mit einer isolierenden Schicht überzogenen Metall, in dessen Innerem sich ein Thermometer (Widerstandsthermometer, Thermoelement, Thermistor) befindet. Auf dem Rohr liegt eine saugfähige Schicht aus Glasfasergewebe, die mit der gesättigten wäßrigen Lithiumchloridlösung getränkt ist. Um die Schicht werden drahtförmige Metallelektroden gewickelt. Schließt man die Elektroden an eine Spannungsquelle an, so wird aufgrund der elektrolytischen Leitfähigkeit der Salzlösung ein Strom durch die Salzlösung fließen, der den Fühler aufheizt. Steigt die Fühlertemperatur über die Gleichgewichtstemperatur T_g, so ist der Dampfdruck der LiCl-Lösung höher als der Dampfdruck des den Fühler umgebenden feuchten Gases. Infolge dieser Dampfdruckdifferenz gibt die LiCl-Lösung Wasser an die Umgebung ab, ein Teil der Lösung kristallisiert aus, und die Leitfähigkeit und damit

auch die Heizleistung des Fühlers nehmen ab. Dadurch kühlt sich der Fühler wieder ab, bis die Fühlertemperatur unter die Gleichgewichtstemperatur gesunken ist. Dann nimmt aber der Fühler wieder Wasser aus der Umgebung auf, das auskristallisierte LiCl wird gelöst, und die Leitfähigkeit und damit auch die Heizleistung des Fühlers steigen wieder an. Nach einigen Regelschwingungen stellt sich schließlich die Gleichgewichtstemperatur von selbst ein.

Leitfähigkeitshygrometer nutzen die Abhängigkeit der elektrischen Leitfähigkeit fester und flüssiger Stoffe von ihrer Feuchte (vgl. 7.3.12.1.) aus. Als Feuchtefühler verwendet man hygroskopische Materialien, deren Feuchte sich auf einen Gleichgewichtswert einstellt, der von der relativen Feuchte der Umgebung abhängt. Bei den elektrolytischen Leitfähigkeitshygrometern dient die elektrische Leitfähigkeit ungesättigter Salzlösungen als Meßgröße. Das empfindliche Element des elektrolytischen Feuchtefühlers ist meist ein flüssiger oder trockener Film des Elektrolyten, der so lange Wasser aus dem umgebenden feuchten Gas aufnimmt, bis der Dampfdruck der Lösung gleich dem Dampfdruck des feuchten Gases ist. Die Konzentration der Lösung hängt damit von der Feuchte des umgebenden Gases ab. Da die Konzentration der Lösung ihre Leitfähigkeit bestimmt, läßt sich auf diese Weise die Gasfeuchte bestimmen. Das feuchteempfindliche Element der Kohleschichthygrometer besteht aus einer dünnen Schicht aus einem hygroskopischen Bindemittel und einer Suspension feiner Kohleteilchen. Die Leitfähigkeit dieser Schicht hängt von ihrer Feuchte ab. Bei den Aluminiumoxidfeuchtefühlern dient eine elektrolytisch auf reines Aluminium aufgebrachte Aluminiumoxidschicht als feuchteempfindliches Element.

Quarzhygrometer beruhen auf der Änderung der Eigenfrequenz eines Schwingquarzes bei der Adsorption von Wasser auf seiner Oberfläche. Zur Erhöhung der Empfindlichkeit wird auf der metallisierten Quarzoberfläche eine dünne Schicht aus einem hygroskopischen Material aufgebracht.

Das Wirkungsprinzip der *coulometrischen Hygrometer* beruht auf einer kontinuierlichen Absorption von Wasserstoff durch Phosphorpentoxid (P_2O_5) mit anschließender elektrolytischer Zersetzung des absorbierten Wassers. Das feuchte Gas strömt durch ein Rohr, in dem sich als feuchteempfindliches Element eine dünne Schicht aus teilweise hydratisiertem P_2O_5 befindet. An dieser Schicht liegen mit einer Gleichspannungsquelle verbundene Elektroden. Bei der Absorption von Wasser bildet sich zunächst nach der Reaktion

$$P_2O_5 + H_2O \rightarrow 2\,HPO_3$$

Metaphosphorsäure, die anschließend unter Rückbildung von P_2O_5 elektrolytisch zerlegt wird,

$$2\,HPO_3 \rightarrow H_2 + \frac{1}{2}O_2 + P_2O_5.$$

Der sich ergebende Elektrolysestrom ist der absoluten Feuchte des Gases und der Durchflußgeschwindigkeit des Gases proportional.

Infrarothygrometer. Wie im Falle der Feuchtemessung fester und flüssiger Stoffe (vgl. 7.3.12.1.) lassen sich auch bei der Gasfeuchtemessung optisch-spektroskopische Methoden einsetzen. Die Infrarothygrometer nutzen die intensiven Absorptionsbanden im Rotationsschwingungsspektrum des Wassers im infraroten Spektralbereich bei Wellenlängen zwischen 0,94 und ≈6,5 µm aus.

Ultravioletthygrometer beruhen auf der Strahlungsabsorption durch Wasserdampf bei der Wellenlänge von 121,56 nm, der sogenannten Lyman-Alpha-Linie des atomaren Wasserstoffs. Bei dieser Wellenlänge hat Wasserdampf einen relativ hohen Absorptionskoeffizienten, so daß eine empfindliche Feuchtemessung möglich ist. Mit den optischen Feuchtemeßverfahren wird die absolute Gasfeuchte ermittelt.

7.4. Messung elektrischer und magnetischer Größen

7.4.1. Messung elektrischer Größen

7.4.1.1. Elektrische Meßgeräte. In den direkt anzeigenden elektrischen Meßgeräten wird die zu messende elektrische Größe (Spannung, Strom) in eine meist mechanische Größe (Kraft, Kraftmoment) umgewandelt und der Meßwert durch Vergleich dieser Größe mit einer entsprechenden Maßverkörperung gebildet. In Abhängigkeit von den für die Umwandlung ausgenutzten physikalischen Gesetzmäßigkeiten unterscheidet man zwischen elektromagnetischen, elektrostatischen und elektrothermischen Meßgeräten.

Elektromagnetische Meßgeräte nutzen die Tatsache aus, daß ein stromdurchflossener Leiter in einem Magnetfeld eine Kraft erfährt, die dem Strom I durch den Leiter und der magnetischen Flußdichte des Magnetfeldes, in dem sich der Leiter befindet, proportional ist. Der am häufigsten anzutreffende Vertreter elektromagnetischer Meßgeräte ist das *Drehspulinstrument* (Abb. 7.4.1–1). Das Magnetfeld wird mit Hilfe des Permanentmagneten, der Weicheisenpolschuhe und des ebenfalls ortsfesten Weicheisenkerns erzeugt, der nur einen schmalen Luftspalt für die Drehung der Spule freiläßt. Der bewegliche Teil des Instruments, das sogenannte bewegliche Organ, besteht aus der Drehspule, der den Zeiger tragenden Achse und dem Zeiger mit den Äquilibriergewichten. Die Ruhelage der Spule, die dem Nullpunkt des Instruments entspricht, wird durch die an

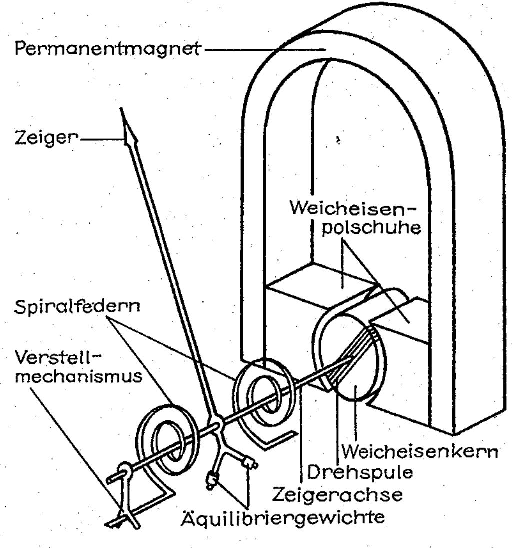


Abb. 7.4.1–1 Aufbau eines Drehspulinstrumentes

der Achse und am Gehäuse befestigten Spiralfedern bestimmt, die gleichzeitig als Stromzuführung dienen. Mit Hilfe eines zusätzlichen Verstellmechanismus ist eine Veränderung der Ruhelage und damit eine Korrektur der Nullpunkteinstellung des Instruments möglich. Fließt nun durch die Spule ein konstanter elektrischer Strom I, so wirken auf die im Magnetfeld befindlichen Leiterstücken der Spule Kräfte, die auf die Spule ein Drehmoment ausüben. Die Spule wird aus ihrer Ruhelage ausgelenkt und führt eine Drehbewegung aus. Die hierbei auftretende Deformation der Spiralfedern verursacht ein Rückstellmoment, welches dieser Drehbewegung entgegenwirkt. Es wird sich schließlich ein Gleichgewichtszustand einstellen, in dem das durch den Strom verursachte Drehmoment der Spule gleich dem Rückstellmoment der Spiralfedern ist. Dem entspricht eine Auslenkung des Zeigers um einen Winkel φ aus seiner Nullpunktstellung, die dem Strom I und der magnetischen Flußdichte B proportional ist, $\varphi = aIB$. Die Proportionalitätskonstante a hängt von der Form und der Windungszahl der Spule und den elastischen Eigenschaften der Spiralfedern ab. Das Drehspulinstrument ist ein Gerät für die Messung von Gleichströmen. Sein Meßbereich erstreckt sich von $I = 0$ bis $I_{\text{end}} = \varphi_{\text{end}}/(aB)$, wobei φ_{end} den maximalen Ausschlagwinkel des Zeigers angibt. Die Messung höherer Ströme als I_{end} ist durch die Verwendung von Stromteilerschaltungen, d. h. durch eine Parallelschaltung von Widerständen zum Meßgerät möglich. Meist wählt man einen ganzen Satz von Parallelwiderständen, die je nach der Größe des zu messenden Stroms I mit Hilfe eines Schalters zugeschaltet werden können (Abb. 7.4.1–2). Für die in der Abb. gezeigte Schalterstellung gilt für den durch das Instrument fließenden Strom I_g die Beziehung $I_g = IR_1/(R_1 + R_g)$, wobei R_g der von den Zuleitungen und der Drahtwicklung der Drehspule herrührende Innenwiderstand des Drehspulinstruments ist. Bei $R_1 = (1/2)R_g$ bedeutet das z. B. eine Erweiterung des Meßbereiches um den Faktor 3, bei $R_1 = (1/9)R_g$ um den Faktor 10.

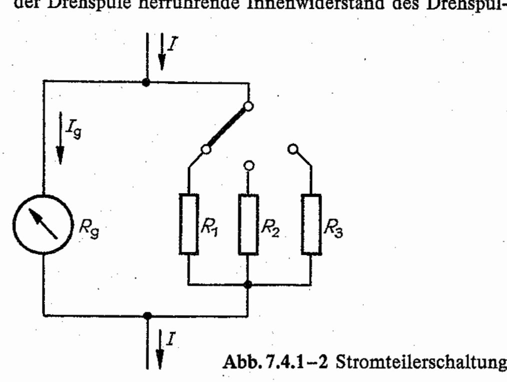


Abb. 7.4.1–2 Stromteilerschaltung

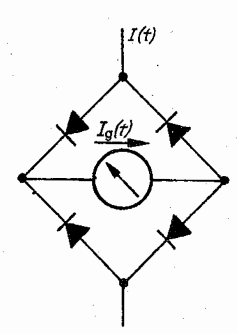


Abb. 7.4.1–3 Gleichrichtwertmessung eines Wechselstromes mit dem Drehspulinstrument

Im Falle sinusförmiger Wechselströme $I(t) = I_0 \sin \omega t$ bekannter Kreisfrequenz ω lassen sich Drehspulinstrumente in Zusammenschaltung mit Gleichrichtern für die Messung der unbekannten Stromamplitude I_0 einsetzen. Eine mögliche Schaltung zeigt Abb. 7.4.1–3. Für diese Schaltung hat der durch das Instrument fließende Strom $I_g(t)$ die Form $I_g(t) = I_0 |\sin \omega t|$. Das Meßinstrument zeigt dann den zeitlichen Mittelwert $\bar{I}$ dieses Stroms an, der die Größe $\bar{I} = (2/\pi)I_0$ hat und als *Gleichrichtwert des Wechselstromes* bezeichnet wird.

Will man mit dem Drehspulinstrument die an einem Verbraucherwiderstand R_v liegende Spannung $U_v = IR_v$ messen, so wird das Meßgerät parallel zum Verbraucher geschaltet. Eine Überlastung des Instruments wird dabei durch Reihenschaltung von Widerständen geeigneter Größe vermieden (Abb. 7.4.1–4). Für die in der Abb. gezeigte Schalterstellung gilt für den durch das Meßgerät fließenden Strom $I_g = U_v/(R_g + R_1)$, d. h. I_g und damit auch der Ausschlag des Drehspulinstrumentes sind der gesuchten Spannung U_v proportional. Die Messung sinusförmiger Wechselspannungen ist mit einer zu Abb. 7.4.1–3 analogen Schaltung möglich.

Elektrodynamisches Instrument. Es unterscheidet sich vom Drehspulinstrument dadurch, daß das Magnetfeld, in dem sich die Drehspule befindet, nicht durch einen Permanentmagneten, sondern durch einen Elektromagneten erzeugt wird. Fließt nun der zu messende Strom nicht nur durch die Drehspule, sondern auch durch die das Magnetfeld erzeugende Feldspule, so ist der Auschlagwinkel des Zeigers des Instruments dem Quadrat des zu messenden Stroms I proportional, $\varphi = bI^2$. Die Konstante b hängt von den Parametern der Drehspule (Form, Windungszahl), den Parametern der Feldspule (Form, Windungszahl, magnetische Permeabilität des verwendeten Weicheisenkerns) und der Geometrie der Anordnung ab. Aus der obigen Beziehung folgt, daß der Ausschlag des elektrodynamischen Instrumentes vom Vorzeichen des Stromes unabhängig ist. Es kann somit unmittelbar für die Messung der Amplitude sinusförmiger Wechselströme eingesetzt werden, wobei der *Effektivwert des Wechselstromes* $I_{\text{eff}} = (1/\sqrt{2})I_0$ gemessen wird. Die Messung von Spannungen mit dem elektrodynamischen Instrument erfolgt in einer zur Abb. 7.4.1–4 analogen Schaltung, eine Meßbereichserweiterung bezüglich Strom- und Spannungsmessung erfolgt in Analogie zum Drehspulinstrument durch Parallelschaltung (vgl. Abb. 7.4.1–2) oder Reihenschaltung (vgl. Abb. 7.4.1–4) von Widerständen. Mit dem elektrodynamischen Instrument kann auch die an einen Verbraucherwiderstand R_v abgegebene Leistung gemessen werden. Eine mögliche Schaltung für die Leistungsmessung zeigt Abb. 7.4.1–5. Die Feldspule mit dem Widerstand R_f wird in Reihe mit dem Verbraucher R_v geschaltet, während die Drehspule mit dem Widerstand R_s und einem der Strombegrenzung dienenden Widerstand R parallel zum Verbraucher und der Feldspule liegt. Der Ausschlag des Meßgerätes ist dann $\varphi = bI_gI_v = bI_vU/(R_g + R)$, d. h., der an den Verbraucher abgegebenen Leistung $P = I_vU$ direkt proportional.

Elektrostatische Meßgeräte basieren auf dem Coulomb-Gesetz, d. h., sie nutzen die Tatsache aus, daß zwischen geladenen Körpern Kräfte wirken. Sie bestehen aus zwei einander gegenüberliegend angeordneten und gegeneinander beweglichen Elektroden, die eine kapazitive Anordnung

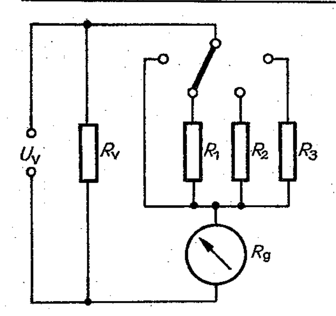


Abb. 7.4.1–4 Spannungsmessung mit dem Drehspulinstrument

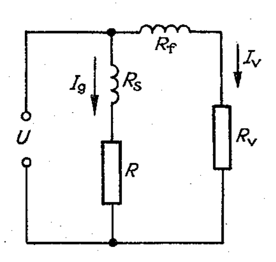


Abb. 7.4.1–5 Schaltung für die Leistungsmessung mit dem elektrodynamischen Instrument

bilden, deren Kapazität C allein durch die Geometrie der Elektroden und durch ihre gegenseitige Lage bestimmt ist (vgl. 8.1.1.2.). Legt man an die Elektroden eine konstante Spannung U an, so werden sie sich entgegengesetzt aufladen, wobei der Betrag Q der Ladung durch $Q = CU$ gegeben ist. Entsprechend dem Coulomb-Gesetz bedeutet das, daß eine Anziehung zwischen den Elektroden auftritt. Praktische Ausführungsformen der elektrostatischen Meßgeräte bestehen meist aus einem Plattenkondensator mit einer feststehenden und einer beweglichen Elektrode. Die Ruhelage der beweglichen Platte, die dem Nullpunkt des Instruments entspricht, kann z. B. durch eine Feder fixiert werden. Wird nun eine Spannung U an den Kondensator gelegt, so wird die auftretende Anziehungskraft zu einer Verringerung des Plattenabstands führen. Im sich einstellenden Gleichgewichtszustand zwischen dieser Anziehungskraft und der Rückstellkraft der die bewegliche Platte haltenden Feder ergibt sich für die Änderung x des Plattenabstandes die Beziehung $x = U^2aC/d$, wobei a eine Konstante ist, die durch die elastischen Eigenschaften der Feder bestimmt wird, C ist die Kapazität des Plattenkondensators und d der Plattenabstand in der Nullage. Die elektrostatischen Meßgeräte sind primär Spannungsmeßgeräte. Da die Auslenkung dem Quadrat der anliegenden Spannung proportional ist, können sie auch ohne zusätzliche Maßnahmen für die Messung des Effektivwertes sinusförmiger Wechselspannungen eingesetzt werden. Strommessungen sind über die Bestimmung des Spannungsabfalls möglich, den der zu messende Strom an einem bekannten Widerstand erzeugt.

Elektrothermische Meßgeräte beruhen auf der Ausnutzung des Jouleschen Gesetzes. Legt man an die Enden eines Leiters, der den elektrischen Widerstand R hat, eine Spannung U an, so wird in ihm ein elektrischer Strom I fließen, dessen Größe durch das Ohmsche Gesetz bestimmt ist, $I = U/R$. Hierbei wird gemäß dem Jouleschen Gesetz die Leistung $P = IU = I^2R = U^2/R$ an den Leiter abgegeben, was zu einer Erwärmung des Leiters führt. Die sich ergebende Temperaturänderung ΔT ist der Leistung P proportional und damit ein Maß für UI, I^2 oder U^2, wenn der Widerstand R als temperaturunabhängig vorausgesetzt werden kann. Von den auf diesem Effekt basierenden Meßgeräten sind heute nur noch die *Thermoumformer* von meßtechnischer Bedeutung. Sie bestehen aus einem Heizer aus einem Widerstandsmaterial mit kleinem Temperaturkoeffizienten in Form eines Drahtes, eines Bandes oder eines Röhrchens. Die durch die zugeführte Leistung bedingte Temperaturerhöhung ΔT wird mit Thermoelementen (vgl. 7.3.9.6.) gemessen. Thermoumformer sind auch für die Strom-, Spannungs- und Leistungsmessung von Wechselströmen einsetzbar.

7.4.1.2. Kompensatoren. Bei Kompensationsmessungen findet stets ein Vergleich der zu messenden Größe mit einer bekannten Größe gleicher Art statt, wobei die Vergleichsgröße stetig regelbar sein muß, so daß die Differenz beider Größen zu Null gemacht werden kann. Das Prinzip eines Gleichspannungskompensators ist in Abb. 7.4.1–6 dargestellt. Eine Normalspannung U_N, die sehr genau bekannt ist, führt zu einem Strom I_N, der durch die Größe des Widerstandes R festgelegt ist. Der Widerstand R ist ein Präzisionspotentiometer (vgl. 8.1.1.1.), welches eine genaue Einstellung von R_1 gestattet. Regelt man nun das Potentiometer so ein, daß der Strom durch das Nullinstrument G verschwindet, so ist der durch den Strom I_N verursachte Spannungsabfall an R_1 gleich der gesuchten Spannung U_x, d. h., es wird $U_x = I_NR_1 = (R_1/R)U_N$.

Der Vorteil der Spannungsmessung mit Kompensatoren besteht vor allem darin, daß die Messung ohne Belastung des Meßobjektes erfolgt und daß die Widerstände der Zuleitungen vom Meßobjekt zum Kompensator unter Abgleichbedingungen (verschwindender Strom durch das Nullinstrument) keine Rolle spielen. Damit hängt die für U_x erzielbare Meßgenauigkeit nur davon ab, wie genau die Widerstände und die Vergleichsspannung U_N bekannt sind.

Wenn die Anforderungen an die Meßgenauigkeit nicht so hoch sind, wird auf eine Normalspannungsquelle U_N hoher Güte verzichtet. Der *Lindeck-Rothe-Kompensator* ist ein Beispiel für einen solchen Gleichspannungskompensator. Die Grundschaltung dieses Kompensators ist in Abb. 7.4.1–7 gezeigt. Die Spannungsquelle U_H erzeugt einen Strom I_H,

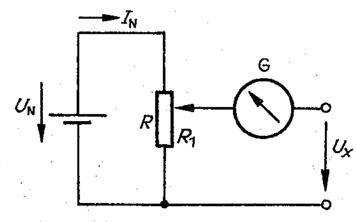


Abb. 7.4.1–6 Prinzip der Gleichspannungskompensation

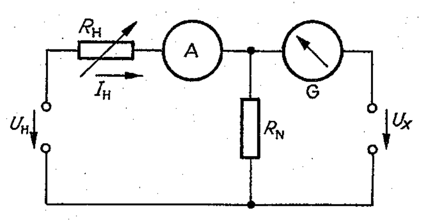


Abb. 7.4.1–7 Lindeck-Rothe-Kompensator

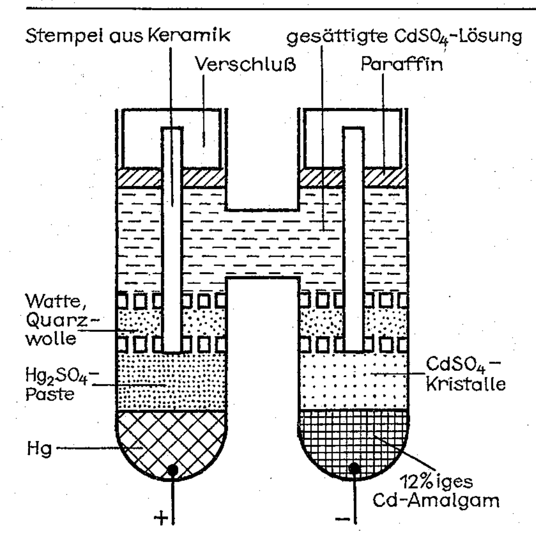


Abb. 7.4.1–8 Weston-Element

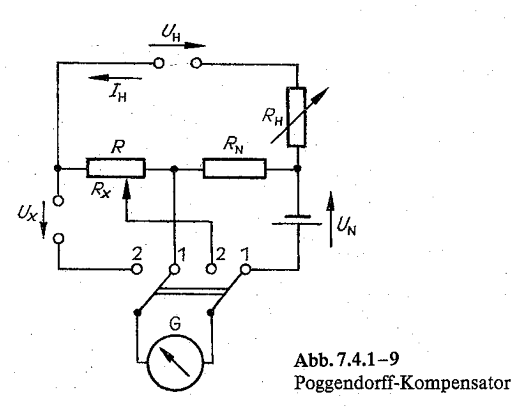


Abb. 7.4.1–9
Poggendorff-Kompensator

der mit einem Präzisionsstrommeßgerät A gemessen werden kann. Mit dem veränderlichen Widerstand R_H wird nun I_H solange verändert, bis der Strom durch das Nullinstrument G verschwindet. Dann erzeugt der Strom I_H an dem genau bekannten Vergleichswiderstand R_N einen Spannungsabfall, der gleich der gesuchten Spannung U_x ist: $U_x = I_H R_N$. Die mit dem Lindeck-Rothe-Kompensator erzielbare Meßgenauigkeit wird durch die Anzeigegenauigkeit des verwendeten Strommeßgerätes A bestimmt.

Bei höheren Ansprüchen an die Meßgenauigkeit wird als Normalspannungsquelle das *Weston-Element* eingesetzt (Abb. 7.4.1–8). Es liefert bei einer Temperatur von 20 °C eine Spannung $U_N = 1{,}018\,65$ V, die sich im Laufe eines Jahres höchstens um $\approx 1\ \mu V$ ändert. Die Temperaturabhängigkeit von U_N folgt im Temperaturbereich von 0 bis 40 °C der Beziehung

$$U_N(\vartheta) = U_N(20\,°C) - 4{,}06 \cdot 10^{-5}(\vartheta - 20) - 9{,}5 \cdot 10^{-7}(\vartheta - 20)^2 + 1 \cdot 10^{-8}(\vartheta - 20)^3,$$

so daß auch eine sehr genaue Korrektur von U_N vorgenommen werden kann, wenn die Meßbedingungen nicht einer Temperatur von $\vartheta = 20$ °C entsprechen. Die Belastbarkeit des Weston-Elements ist gering und kann über längere Zeit $\approx 10\ \mu A$, kurzzeitig maximal 100 μA betragen. Das Weston-Element kann folglich nicht in Schaltungen wie in Abb. 7.4.1–6 oder 7.4.1–7 als Vergleichsspannungsquelle eingesetzt werden. Um trotzdem die mit dem Weston-Element als Normalspannungsquelle erreichbare Meßgenauigkeit zu nutzen, sind Gleichspannungskompensatoren mit sogenannter doppelter Kompensation entwickelt worden, die neben dem Weston-Element noch eine Hilfsspannungsquelle U_H enthalten. Ein Beispiel hierfür ist der häufig verwendete *Poggendorff-Kompensator* (Abb. 7.4.1–9). Der Schalter wird zunächst in Stellung 1 gebracht, so daß das Nullinstrument G in dem aus U_N und einem Präzisionswiderstand R_N gebildeten Kreis liegt. Dann wird der Strom I_H mit Hilfe des veränderlichen Widerstandes R_H so eingeregelt, daß der Strom durch das Nullinstrument G verschwindet, so daß $I_H R_N = U_N$ ist. Anschließend wird der Schalter in Stellung 2 gebracht, d. h., das Nullinstrument G liegt jetzt in dem aus der zu messenden Spannung U_x und dem Widerstand R_x gebildeten Kreis. Der nun erfolgende Abgleich mit Hilfe des Präzisionspotentiometers R ergibt bei verschwindendem Strom durch das Nullinstrument G für die gesuchte Spannung $U_x = I_H R_x = (R_x/R_N) U_N$. Der Hilfsstrom I_H ändert sich beim zweiten Abgleichvorgang nicht, da die ihn bestimmenden Widerstände R_H, R_N und R dabei unverändert bleiben.

Andere Präzisionskompensatoren unterscheiden sich von der Poggendorff-Schaltung nur darin, daß unterschiedliche Widerstandsschaltungen verwendet werden.

Die für genaue Spannungsmessungen mit Kompensatoren erforderlichen Präzisionswiderstände und -potentiometer werden aus Manganindraht, im Falle sehr kleiner Widerstände aus Manganinblech gefertigt. Manganin ist eine Legierung aus 84 % Kupfer, 12 % Mangan und 4 % Nickel, die sich durch einen sehr kleinen Temperaturkoeffizienten des elektrischen Widerstandes auszeichnet. Durch künstliche Alterung wird erreicht, daß die jährliche Änderung der Widerstände unterhalb 1 μΩ liegt. Um eine genaue Einstellung und Ablesung der Widerstandswerte zu ermöglichen, beste-

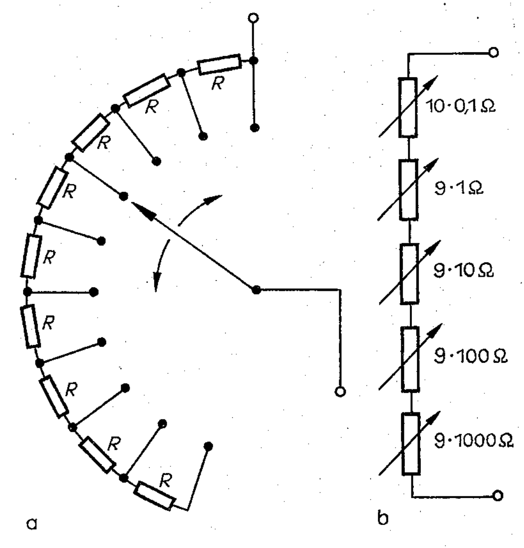


Abb. 7.4.1–10 (a) stufenweise regelbarer Widerstand und (b) Dekadensatz von stufenweise regelbaren Widerständen

hen die Widerstände aus Dekadensätzen von stufenweise regelbaren Widerständen (Abb. 7.4.1–10). Mit dem in Abb. 7.4.1–10b dargestellten Dekadensatz kann man beispielsweise alle Widerstände im Bereich von 0,1 bis 10 000 Ω mit einer Genauigkeit von 0,1 Ω einstellen.
Kompensatoren für die Wechselspannungsmessung sind ebenfalls entwickelt worden, haben aber für Präzisionsmessungen keine Bedeutung erlangt. Das hat zwei Gründe. Erstens ist der Abgleich einer Wechselspannung $U = U_0 \sin(\omega t + \varphi)$ wesentlich komplizierter, da sie i. allg. durch drei Größen bestimmt wird: die Amplitude U_0, die Frequenz ω und die Phase φ. Um also eine Wechselspannung kompensieren zu können, muß die Vergleichsspannung so beschaffen sein, daß alle drei Größen unabhängig voneinander variiert werden können. Zweitens gibt es keine bezüglich ihrer Genauigkeit mit dem Weston-Element vergleichbare Normalquelle für Wechselspannungen.
Aus den in Abb. 7.4.1–7 und 7.4.1–9 angegebenen Grundschaltungen für Gleichspannungskompensatoren lassen sich durch geringfügige Veränderungen die Grundschaltungen für Gleichstromkompensatoren herleiten. Die Schaltung für einen aus dem Lindeck-Rothe-Kompensator (vgl. Abb. 7.4.1–7) abgeleiteten Gleichstromkompensator zeigt Abb. 7.4.1–11. Der Abgleich (verschwindender Strom durch das Nullinstrument G) erfolgt durch Änderung von I_H mit Hilfe des regelbaren Widerstandes R_H. Dem abgeglichenen Zustand entspricht ein gleichgroßer Spannungsabfall an R_N und R_v, d. h., es gilt dann $(I_H - I_x)R_N = I_x R_v$. Hieraus folgt für den zu messenden Strom I_x die Beziehung $I_x = I_H R_N/(R_N + R_v)$. Mit dieser Schaltung wird der Strom ohne Spannungsabfall am Eingang des Meßgerätes gemessen, d. h., die angegebene Kompensationsschaltung hat im abgeglichenen Zustand den für Strommessungen idealen Eingangswiderstand Null. Analog zu dem in Abb. 7.4.1–11 gezeigten Fall läßt sich auch der Poggendorff-Kompensator (vgl. Abb. 7.4.1–9) in einen Gleichstromkompensator umwandeln.

7.4.1.3. Meßbrücken sind Geräte für die Messungen von Ohmschen Widerständen, Kapazitäten und Induktivitäten. Die einfachste Variante einer *Widerstandsmeßbrücke* ist die *Wheatstone-Brücke*, deren Grundschaltung in Abb. 7.4.1–12 dargestellt ist. Sie besteht aus zwei Festwiderständen R_1 und R_2, einem regelbaren Widerstand R_3 und einem Nullinstrument G. R_x ist der unbekannte Widerstand, dessen Größe ermittelt werden soll. Legt man jetzt entsprechend Abb. 7.4.1–12. eine Spannung U an die Schaltung, so wird i. allg. ein Strom durch das Nullinstrument G fließen. Stellt man nun den Widerstand R_3 so ein, daß der Strom durch das Nullinstrument G verschwindet, so bedeutet das, daß die Punkte 1 und 2 auf gleichem Potential liegen, d. h., die Teilerverhältnisse der aus den Widerständen R_1 und R_x einerseits und aus den Widerständen R_2 und R_3 andererseits gebildeten Spannungsteiler müssen einander gleich sein: $R_1/R_x = R_2/R_3$. Hieraus ergibt sich für den gesuchten Widerstand $R_x = (R_1/R_2)R_3$. In Meßbrücken für genaue Messungen besteht der Widerstand R_3, mit dessen Hilfe der Abgleich vorgenommen wird, aus Dekadensätzen von stufenweise regelbaren Widerständen (vgl. 7.4.1.2., Abb. 7.4.1–10). Bei nicht zu hohen Ansprüchen an die Meßgenauigkeit kann für R_3 ein Schiebewiderstand (vgl. 8.1.1.1.) verwendet werden.
Die untere Meßgrenze der Wheatstone-Brücke wird erreicht, wenn die Zuleitungswiderstände zum Meßobjekt R_x und die Übergangswiderstände an den Kontakten zum Anschluß von R_x dieselbe Größenordnung haben wie R_x

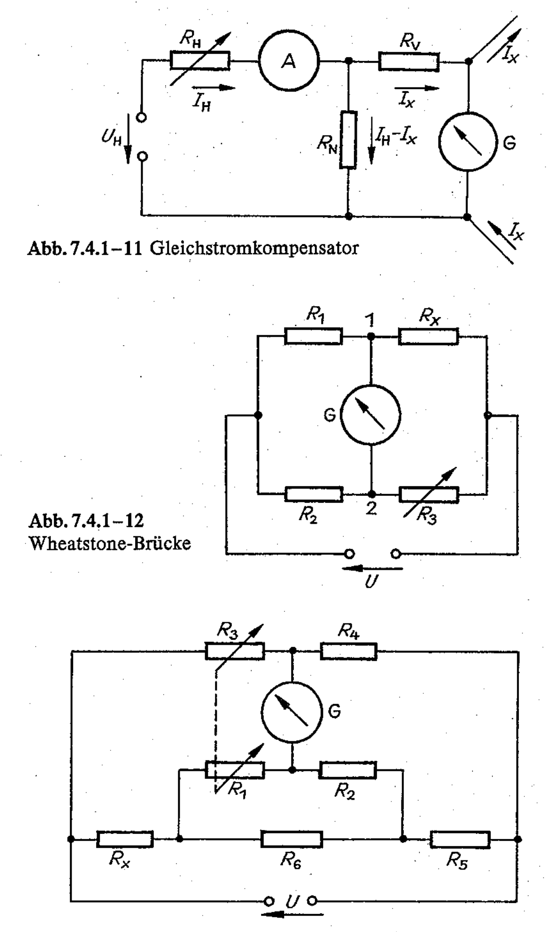


Abb. 7.4.1–11 Gleichstromkompensator

Abb. 7.4.1–12 Wheatstone-Brücke

Abb. 7.4.1–13 Thomson-Brücke

selbst. Das bedeutet, daß Wheatstone-Brücken nur für die Messung von Widerständen oberhalb ≈1 Ω geeignet sind. Die Messung von Widerständen unterhalb ≈1 Ω erfolgt mit der *Thomson-Brücke*, deren Prinzipschaltbild in Abb. 7.4.1–13 gezeigt ist. Hier ist R_x wieder der zu messende Widerstand, R_5 ein bekannter Normalwiderstand gleicher Größenordnung und R_6 der Widerstand der Verbindungsleitung zwischen R_x und R_5. Zur Erzielung einer hinreichend hohen Empfindlichkeit fließt durch diese Widerstände ein recht hoher Strom, der in praktischen Ausführungsformen der Thomson-Brücke bis zu 200 A betragen kann. Die Widerstände R_1, R_2, R_3 und R_4 sind hochohmig und mechanisch so miteinander gekoppelt, daß bei der Einstellung des Abgleiches (verschwindender Strom durch das Nullinstrument G) durch Regelung von R_1 und R_3 stets die Bedingung $R_1/R_2 = R_3/R_4$ erfüllt bleibt. Als Abgleichbedingung ergibt sich dann $R_x = (R_1/R_2)R_5 = (R_3/R_4)R_5$.
Die Thomson-Brücke kann auch durch Regelung von R_5 bei festgehaltenem Verhältnis $R_1/R_2 = R_3/R_4$ abgeglichen werden. Mit der Thomson-Brücke lassen sich Widerstände im Bereich von ≈1 Ω bis herunter zu ≈1 μΩ messen.

Zur Messung von Kapazitäten und Induktivitäten wird die Tatsache ausgenutzt, daß sie in Wechselstromkreisen als komplexe Widerstände wirken (vgl. 8.1.1.2. und 8.1.1.3.). Im Falle einer Kapazität der Größe C ist der entsprechende Widerstand $R_C = 1/(i\omega C)$, im Falle einer Induktivität der Größe L ist $R_L = i\omega L$, wenn ω die Kreisfrequenz des Wechselstromes ist. Kapazitäts- und Induktivitätsmeßbrücken haben vom Prinzip her denselben Schaltungsaufbau wie die Wheatstone-Brücke (vgl. Abb. 7.4.1–12), wobei jedoch jetzt die Ohmschen Widerstände R_1, R_2, R_3 und R_x durch komplexe Widerstände Z_1, Z_2, Z_3 und Z_x zu ersetzen sind und die Brücke nicht mehr mit einer Gleichspannung U, sondern mit einer Wechselspannung $U(t) = U_0 \sin \omega t$ mit bekannter Frequenz ω betrieben werden muß. In vollständiger Analogie zur Wheatstone-Brücke erhält man für den Abgleich der Brücke (verschwindender Wechselstrom durch das Nullinstrument G) die Bedingung $Z_x Z_2 = Z_1 Z_3$. Da die Widerstände Z_i jedoch komplex sind, bedeutet vollständiger Abgleich, daß die Realteile und Imaginärteile beider Seiten der Abgleichbeziehung einander gleich sein müssen,

$$\mathrm{Re}(Z_x Z_2) = \mathrm{Re}(Z_1 Z_3), \mathrm{Im}(Z_x Z_2) = \mathrm{Im}(Z_1 Z_3).$$

Da man jeden komplexen Widerstand auch in der Form $Z = |Z| e^{i\varphi}$ schreiben kann, ist auch ein Abgleich nach Betrag und Phase möglich:

$$|Z_x Z_2| = |Z_1 Z_3|, \varphi_x + \varphi_2 = \varphi_1 + \varphi_3.$$

Daraus folgt, daß jede Kapazitäts- oder Induktivitätsmeßbrücke zwei voneinander unabhängige Abgleichorgane enthalten muß, mit denen nach Real- und Imaginärteil bzw. nach Betrag und Phase abgeglichen werden kann. Für die Feststellung des Abgleichs, also des Zustands verschwindenden Stroms durch das Nullinstrument, kann jedes hinreichend empfindliche Wechselstrommeßgerät verwendet werden.

Die *Wien-Robinson-Brücke* ist ein Beispiel für eine *Kapazitätsmeßbrücke* (Abb. 7.4.1–14). Für die beiden Abgleichbedingungen erhält man $C_1 C_x = 1/(\omega^2 R_1 R_4)$ und $C_1/C_x = R_3/R_2 - R_4/R_1$, woraus sich für die gesuchte Kapazität C_x die Gleichung

$$C_x = C_1 \frac{R_3}{R_2} \frac{1 + (\omega R_1 C_1)^2}{(\omega R_1 C_1)^2}$$

ergibt. Der Abgleich ist frequenzabhängig. Als Abgleichelemente kann man, wie in Abb. 7.4.1–14 angeben, die Widerstände R_1 und R_2 wählen. Eine häufig verwendete *Induktivitätsmeßbrücke* ist die *Maxwell-Wien-Brücke* (Abb. 7.4.1–15). Die beiden Abgleichbedingungen lauten hier $L_x = R_1 R_3 C_2$ und $R_1/R_4 = R_2/R_3$. Im Gegensatz zur Wien-Robinson-Brücke ist hier der Abgleich frequenzunabhängig. Als Abgleichelemente können die Widerstände R_2 und R_3 gewählt werden.

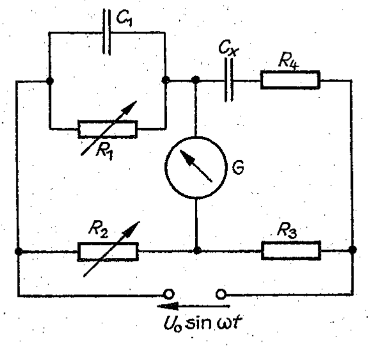


Abb. 7.4.1–14 Wien-Robinson-Brücke

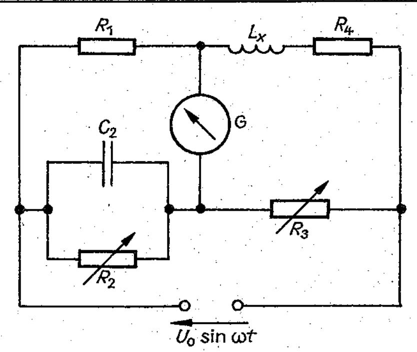


Abb. 7.4.1–15 Maxwell-Wien-Brücke

Wie man aus den angegebenen beiden Beispielen sieht, reicht es aus, einen der Widerstände Z_1, Z_2 und Z_3 der Brücke komplex zu wählen. Unter dieser Bedingung ist es stets möglich, den Abgleich mit Hilfe zweier regelbarer Ohmscher Widerstände einzustellen. Das ist insofern von Vorteil, als man über regelbare Präzisionswiderstände hoher Genauigkeit verfügt (vgl. 7.4.1.2.). Es gibt eine ganze Reihe weiterer Schaltungen für Kapazitäts- und Induktivitätsmeßbrücken, die sich von den hier beschriebenen dadurch unterscheiden, daß meist zwei der Widerstände Z_1, Z_2 und Z_3 komplex sind. Dann kommt man in der Regel nicht umhin, regelbare Kondensatoren als Abgleichelemente vorzusehen. Die mit stufenweise regelbaren Kondensatordekaden erreichbare Genauigkeit liegt jedoch wesentlich unter der Genauigkeit, die mit regelbaren Präzisionswiderständen erzielt werden kann.

7.4.1.4. Verstärker. Durch den Einsatz von Verstärkern in der Meßtechnik wird es möglich, kleinste Spannungen und Ströme zu messen und auch das Meßsystem an den Meßgrößenaufnehmer anzupassen. An in der Meßtechnik zu verwendende Verstärker sind die folgenden Anforderungen zu stellen:

- hinreichend große und frequenzunabhängige Verstärkung in einem vorgegebenen Frequenzbereich,
- linearer Zusammenhang zwischen Eingangs- und Ausgangsgröße,
- keine Rückwirkung auf die Signalquelle,
- niedriger Ausgangswiderstand,
- kleine Nullpunkt- und Verstärkungsdrift in Abhängigkeit von der Temperatur und der Meßdauer,
- die im Verstärker selbst erzeugten Störspannungen (Brumm- und Rauschspannungen) sollen klein sein.

Diese Forderungen werden weitgehend von integrierten Operationsverstärkern (vgl. 8.2.3.1.) erfüllt. Sie dienen deshalb als Grundbausteine für die meisten in der Meßtechnik eingesetzten Verstärkerschaltungen. Durch entsprechende äußere Beschaltung ist es möglich, das Übertragungsverhalten des Operationsverstärkers den jeweiligen Erfordernissen anzupassen (vgl. 8.2.3.1.).

Für die Abschätzung der Einsatzmöglichkeiten eines Verstärkers für meßtechnische Zwecke ist nicht nur die Kenntnis des Betrages der Verstärkung, sondern auch ihrer Frequenzabhängigkeit von Interesse. Sie wird durch den *Frequenzgang* des Verstärkers beschrieben, der i. allg. experi-

mentell bestimmt wird. Für seine Messung gibt man ein Testsignal der Form $x_e(t) = x_{e0} \exp(i\omega t)$ auf den Eingang des Verstärkers. Das Ausgangssignal wird dann ebenfalls die Frequenz ω haben, jedoch i. allg. eine andere Amplitude haben und eine Phasenverschiebung φ gegenüber dem Eingangssignal zeigen, so daß für das Ausgangssignal $x_a(t) = x_{a0} \exp[i(\omega t + \varphi)]$ zu schreiben ist. Der Frequenzgang $G(i\omega)$ ist dann durch

$$G(i\omega) = \frac{x_a(t)}{x_e(t)} = \frac{x_{a0}}{x_{e0}} \exp(i\varphi)$$

definiert. Die Frequenzabhängigkeit des Betrages $|G(i\omega)| = |x_{a0}/x_{e0}|$ des Frequenzganges wird als *Amplitudengang*, die Frequenzabhängigkeit $\varphi(i\omega)$ des Phasenwinkels als *Phasengang* bezeichnet. Jeder Verstärker wird nur innerhalb eines bestimmten Frequenzbereiches zwischen einer unteren und oberen *Grenzfrequenz* $f_{gu} = \omega_{gu}/2\pi$ bzw. $f_{g0} = \omega_{g0}/2\pi$ eine frequenzunabhängige Verstärkung aufweisen, während die Verstärkung für $f < f_{gu}$ und $f > f_{g0}$ mehr oder weniger steil abfällt. Diese Grenzfrequenzen sind so definiert, daß der Amplitudengang bei $\omega = \omega_{gu}$ bzw. ω_{g0} jeweils auf das $1/\sqrt{2}$fache seines Maximalwertes abgesunken ist. Die Differenz $B = f_{g0} - f_{gu}$ wird als *Bandbreite* des Verstärkers bezeichnet. Neben dem Frequenzgang ist das Einstellverhalten eines Verstärkers für seine Einsatzmöglichkeiten von Bedeutung. Führt man dem Verstärkereingang ein zu einem bestimmten Zeitpunkt sprungartig einsetzendes, dann aber zeitlich konstantes Signal x_{e0} zu, so wird sich das entsprechende Ausgangssignal x_{a0} mit einer gewissen zeitlichen Verzögerung einstellen. Dieses Verhalten des Meßsystems wird durch die *Übergangsfunktion* $h(t)$ beschrieben, die durch die Beziehung $h(t) = x_a(t)/x_{e0}$ definiert ist. In Analogie zu den Grenzfrequenzen, die für die Charakterisierung des Frequenzganges eingeführt wurden, läßt sich hier eine charakteristische Zeit t_E definieren, die als *Einschwingzeit* bezeichnet wird. Dabei versteht man unter t_E die Zeit, die nach dem Einsetzen des Sprungsignals am Verstärkereingang wenigstens vergehen muß, damit die Übergangsfunktion $h(t)$ innerhalb der Grenzen von ± 5 % ihres Endwertes $h_0 = x_{a0}/x_{e0}$ bleibt. Zwischen der so definierten Einschwingzeit und der oberen Grenzfrequenz f_{g0} des Verstärkers besteht die Beziehung $f_{g0} = 1/2t_E$, die als *Abtasttheorem* bezeichnet wird. Mitunter wird anstelle der Einschwingzeit t_E eine Anstiegszeit t_A zur Charakterisierung der Übergangsfunktion angegeben. Sie gibt die Zeitdauer an, in der die Funktion $h(t)$ von 10 % auf 90 % ihres Endwertes h_0 ansteigt. Näherungsweise gilt $t_A \approx 0{,}7 t_E$.

Operationsverstärker haben eine untere Grenzfrequenz $f_{gu} = 0$ und je nach Typ Bandbreiten (d. h., obere Grenzfrequenzen) bis zu 1 MHz und mehr. Sie sind damit gut für den Aufbau von Breitbandverstärkern geeignet.

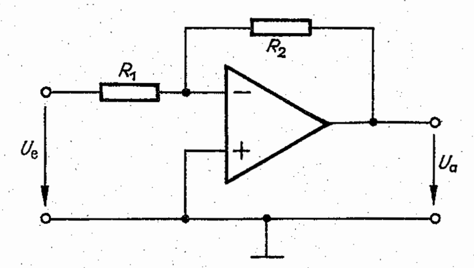


Abb. 7.4.1–16 Invertierender Verstärker mit Gegenkopplung

Breitbandverstärker. Abb. 7.4.1–16 zeigt die Grundschaltung für einen Breitbandverstärker mit Spannungseingang. Es handelt sich um einen invertierenden Verstärker mit Gegenkopplung, dessen Spannungsverstärkung sich zu $v_u = -R_2/R_1$ ergibt, wenn man von der Annahme ausgeht, daß der Operationsverstärker als idealer Operationsverstärker (vgl. 8.2.3.1.) angesehen werden kann. Diese Schaltung bietet den Vorteil, daß die Spannungsverstärker in sehr einfacher Weise durch geeignete Wahl der Widerstände R_1 und R_2 eingestellt werden kann. Für den idealen Operationsverstärker ist der Eingangswiderstand R_{ein} dieses Verstärkers gleich $R_{ein} = R_1$. Hierin kann ein Nachteil dieser Verstärkerschaltung bestehen, da R_1 im Interesse einer hohen Verstärkung nicht beliebig hoch gewählt werden kann, andererseits aber für eine Spannungsmessung ein möglichst hoher Eingangswiderstand des Verstärkers zu fordern ist.

Der Nachteil des niedrigen Eingangswiderstandes der Schaltung in Abb. 7.4.1–16 läßt sich durch eine Schaltung beheben, wie sie für den einfachsten Fall in Abb. 7.4.1–17a dargestellt ist. Hier wird die an den nichtinvertierenden

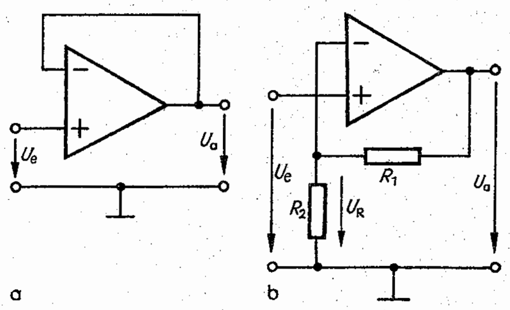


Abb. 7.4.1–17 Elektrometerschaltungen mit einem Operationsverstärker

Eingang gelegte Eingangsspannung U_e durch die Rückführung der Ausgangsspannung U_a auf den invertierenden Eingang wieder kompensiert. Aus diesem Grunde bezeichnet man diese Schaltungsart als Kompensations- oder Elektrometerschaltung. Für einen idealen Operationsverstärker ist die Spannungsverstärkung $v_u = 1$, der Eingangswiderstand R_{ein} unendlich groß und der Ausgangswiderstand $R_{aus} = 0$. Für einen realen Operationsverstärker, bei dem endliche Werte für die Differenzspannungsverstärkung, den Differenzeingangswiderstand und den Gleichtakteingangswiderstand zu berücksichtigen sind (vgl. 8.2.3.1.), lassen sich mit dieser Schaltung Eingangswiderstände $R_{ein} \gtrsim 10^{12}\,\Omega$ und Ausgangswiderstände $R_{aus} \lesssim 10^{-2}\,\Omega$ erzielen, wenn der Operationsverstärker in seinen Eingangsstufen mit Feldeffekttransistoren bestückt ist. Hieraus folgt, daß das Hauptanwendungsgebiet dieser Schaltung in der Impedanzwandlung liegt, um hochohmige Signalquellen an die weitere Meßeinrichtung, z. B. einen Verstärker mit einer Schaltung nach Abb. 7.4.1–16 anzupassen.

Eine gegenüber Abb. 7.4.1–17a abgewandelte Schaltung zeigt Abb. 7.4.1–17b. Hier wird nicht mehr die gesamte Ausgangsspannung U_a, sondern über den aus zwei Widerständen R_1 und R_2 bestehenden Spannungsteiler nur ein Teil $U_R = U_a R_2/(R_1 + R_2)$ der Ausgangsspannung zurückgekoppelt. Für einen idealen Operationsverstärker erhält man für diesen Fall eine Spannungsverstärkung $v_u = 1 + R_1/R_2$ und wieder einen unendlich hohen Eingangswiderstand R_{ein} und einen Ausgangswiderstand $R_{aus} = 0$. Betrachtet man

einen realen Operationsverstärker, so zeigt sich, daß der Eingangswiderstand der Schaltung in Abb. 7.4.1–17b kleiner ist als im Falle der Schaltung in Abb. 7.4.1–17a. Dafür hat die Verstärkerschaltung nach Abb. 7.4.1–17b den Vorteil, daß man neben der Spannungsverstärkung v_u auch den Eingangswiderstand R_{ein} über eine entsprechende Dimensionierung von R_1 und R_2 zumindest innerhalb gewisser Grenzen regeln kann.
Steht man vor der Aufgabe, sehr kleine Ströme zu messen, so muß man über Verstärker mit sehr niedrigem Eingangswiderstand R_{ein} verfügen. Eine einfache Schaltung für diesen Anwendungszweck läßt sich aus Abb. 7.4.1–16 ableiten und ist in Abb. 7.4.1–18 dargestellt. Für einen idealen Operationsverstärker ergibt sich hier für den Zusammenhang zwischen dem Eingangsstrom I_e und der Ausgangsspannung U_a die einfache Beziehung $U_a/I_e = -R_1$. Die Schaltung in Abb. 7.4.1–18 wirkt somit als Strom-Spannungs-Umsetzer, die Größe U_a/I_e wird als Übertragungsfaktor der Schaltung bezeichnet. Für einen idealen Operationsverstärker ist der Eingangswiderstand der in Abb. 7.4.1–18 gezeigten Schaltung gleich $R_{ein} = 0$, während er im Falle eines realen Operationsverstärkers den Wert $R_{ein} = R_1/v_D$ annimmt. Hier ist v_D die Differenzspannungsverstärkung des Operationsverstärkers (vgl. 8.2.3.1.). Man kann also mit der Schaltung in Abb. 7.4.1–18 sehr niedrige Eingangswiderstände erzielen, die sich durch geeignete Wahl von R_1 der jeweiligen Aufgabenstellung anpassen lassen.

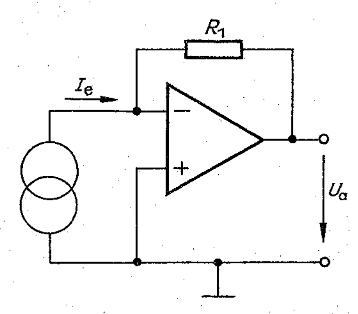


Abb. 7.4.1–18 Strom-Spannungs-Wandler

Die beschriebenen Schaltungen (Bild 7.4.1–16 bis 18) bilden die Grundschaltungen aller Breitbandverstärker für meßtechnische Zwecke. Sie können durch Variationen in der Beschaltung der Operationsverstärker den konkreten Problemstellungen angepaßt werden. Die ebenfalls mit Hilfe von Operationsverstärkern realisierbaren Strom-Strom- und Spannungs-Strom-Wandler spielen für Meßverstärker kaum eine Rolle, da insbesondere in der digitalen Meßtechnik die Messung aller elektrischen Größen stets auf die Messung einer Spannung zurückgeführt wird (vgl. 7.4.1.7.).

Selektive Verstärker. Bei einem Breitbandverstärker besteht die Zielstellung darin, Signale in einem möglichst breiten Frequenzband mit konstanter Amplitude und Phase zu übertragen. Mitunter kann es jedoch bei bestimmten Aufgabenstellungen oder unter bestimmten Meßbedingungen erwünscht sein, nur Signale mit einer bestimmten Frequenz oder innerhalb eines gewissen Frequenzbereichs zu verstärken. Es kann auch vorkommen, daß Signale in einem bestimmten Frequenzbereich unterdrückt werden sollen. Das ist insbesondere dann von Bedeutung, wenn in diesem Frequenzbereich unerwünschte Störsignale auftreten. Die

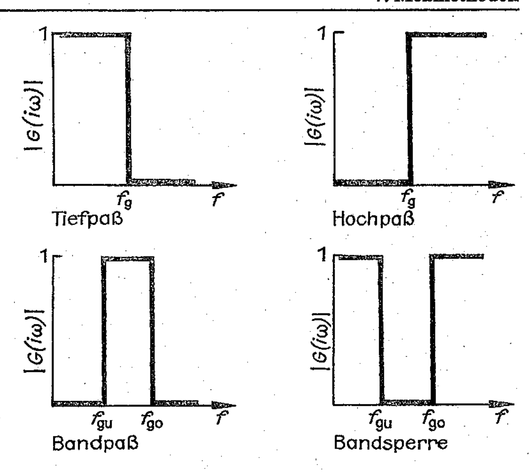


Abb. 7.4.1–19 Idealer Amplitudengang der vier verschiedenen Filtertypen

Lösung dieser Aufgabenstellung erfolgt mit *selektiven Verstärkern.* Solche Verstärker lassen sich aus den beschriebenen Breitbandverstärkern aufbauen, indem man in das Verstärkersystem zusätzliche Filterschaltungen einbaut, die entweder Signale in einem bestimmten Frequenzbereich unterdrücken oder Signale nur in einem bestimmten Frequenzbereich hindurchlassen. Nach ihren Eigenschaften unterscheidet man vier Arten von Filterschaltungen: Tiefpaß, Hochpaß, Bandpaß und Bandsperre. Im stets anzustrebenden Idealfall sind diese Filterschaltungen durch den in Abb. 7.4.1–19 dargestellten Amplitudengang $|G(i\omega)|$ charakterisiert. In der Praxis lassen sich diese idealen Amplitudengänge niemals vollständig realisieren. Alle Filterschaltungen zeigen anstelle des sprungartigen Übergangs zwischen Durchlaß- und Sperrbereichen ein mehr oder weniger ausgedehntes Übergangsgebiet in der Umgebung der Grenzfrequenzen, welches durch eine endliche Flankensteilheit des Amplitudengangs an den Grenzfrequenzen und durch einen frequenzabhängigen Amplituden- und Phasengang in der Umgebung der Grenzfrequenzen gekennzeichnet ist.
Man unterscheidet zwischen passiven und aktiven Filtern.
Passive Filter sind frequenzabhängige Spannungsteilerschaltungen, die neben Ohmschen Widerständen noch Kapazitäten oder Induktivitäten oder beide, d.h., frequenzabhängige Widerstände enthalten (vgl. 8.1.1.2. und 8.1.1.3.). Die einfachsten Schaltungsvarianten, die sich für die vier Filtertypen durch Kombination von Ohmschen Widerständen und Kapazitäten realisieren lassen, sind in Abb. 7.4.1–20 dargestellt. Die Amplitudengänge dieser einfachen Schaltungen weichen jedoch noch beträchtlich von den idealen Amplitudengängen in Abb. 7.4.1–19 ab. So gilt für den in Abb. 7.4.1–20a dargestellten Tiefpaß $|G(i\omega)| = [1 + (\omega RC)^2]^{-1/2}$, die Grenzfrequenz f_g dieses Tiefpasses liegt bei $f_g = 1/2\pi RC$. Für Frequenzen oberhalb f_g fällt der Amplitudengang nur langsam ab und hat bei $f = 3f_g$ noch einen Wert von $\approx 0{,}32$. Die Filtereigenschaften lassen sich verbessern, wenn man mehrere Tiefpässe vom in Abb. 7.4.1–20a dargestellten Typ in Reihe schaltet. Dasselbe gilt für die anderen in Abb. 7.4.1–20 angegebenen Filterschaltungen.
Aktive Filter basieren auf der Ausnutzung von Verstärkern,

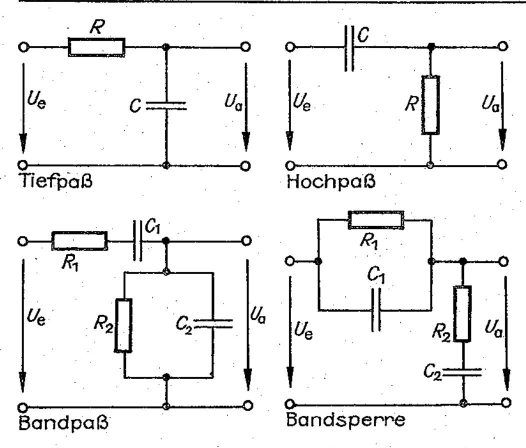


Abb. 7.4.1–20 RC-Schaltungen für (a) Tiefpaß, (b) Hochpaß, (c) Bandpaß und (d) Bandsperre

die in ihrer Beschaltung frequenzabhängige Widerstände, in der Regel Kapazitäten enthalten. Dadurch wird eine Frequenzabhängigkeit der Spannungsverstärkung v_u erreicht. Abb. 7.4.1–21 zeigt ein Beispiel für ein aktives Filter. Unter Annahme eines idealen Operationsverstärkers erhält man für den Betrag der Verstärkung dieser Schaltung

$$|v_u| = [(1 - \omega^2 C_1 C_2 R^2)^2 + 9(\omega C_1 R)^2]^{-1/2},$$

d. h., diese Schaltung hat die Eigenschaften eines Tiefpasses.

Die in selektive Verstärker eingebauten Filterschaltungen sind meist in Stufen oder auch stufenlos regelbar, so daß der Verstärker in verschiedenen Frequenzbereichen eingesetzt werden kann.

Modulationsverstärker. Die bisher beschriebenen direktgekoppelten Verstärker haben nur eine begrenzte Empfindlichkeit, d. h., es können nur Eingangssignale oberhalb einer gewissen Mindestgröße sicher nachgewiesen und gemessen werden. Die Ursache hierfür liegt in der temperatur- und zeitabhängigen Nullpunktdrift des Verstärkers, im Rauschen des Verstärkers und im möglicherweise vorhandenen Einfluß von äußeren Fremdsignalen auf das Meßsystem. Diese Störgrößen führen zur Anzeige einer Ausgangsspannung am Verstärker, die nicht durch eine entsprechende Änderung der Eingangsspannung bedingt ist. Ist diese durch die Störgrößen verursachte Ausgangsspannung von der gleichen Größe oder größer als die durch das zu messende Eingangssignal erzeugte Ausgangsspannung, so wird eine eindeutige Trennung des Meßsignals von den Störgrößen sehr erschwert oder überhaupt nicht mehr möglich. Die Nullpunktdrift eines Verstärkers kann durch geeignete Wahl der verwendeten Bauelemente und der Verstärkerschaltung weitgehend unterdrückt werden, so daß sie als Störgröße nur eine untergeordnete Rolle spielt. Das Rauschen dagegen kann prinzipiell niemals vollständig vermieden werden. Während die Rauschleistung bei Frequenzen oberhalb ≈100 Hz nahezu konstant ist, steigt sie nach niedrigeren Frequenzen bedingt durch das Funkelrauschen stark an. Die Rauschleistung P_R wächst dabei etwa mit $P_R \sim 1/f$. Das macht sich insbesondere bei der Messung sehr kleiner Gleichspannungen nachteilig bemerkbar und begrenzt die Empfindlichkeit von Gleichspannungsverstärkern. Ausgehend von der Frequenzabhängigkeit der Rauschleistung bietet sich folgende Möglichkeit für die Messung kleiner Eingangssignale an. Die zu messende Gleichspannung wird mit Hilfe eines Modulators in einen Bereich höherer Frequenz transformiert, in dem die Rauschleistung bereits auf ihren nahezu konstanten Wert abgesunken ist und in dem auch keine zusätzlichen Störungen durch Fremdeinflüsse auftreten. Dieses höherfrequente Signal wird mit Hilfe eines selektiven Wechselspannungsverstärkers verstärkt, wodurch gleichzeitig alle nieder- und höherfrequenten Signalanteile unterdrückt werden. Schließlich wird dieses verstärkte Signal mit Hilfe eines Demodulators wieder in eine Gleichspannung umgewandelt, wobei das im Modulationsfrequenzgebiet vorhandene Rauschen einschließlich des Modulationsfrequenzsignals selbst noch durch einen Tiefpaß unterdrückt wird. Verstärkersysteme, die nach diesem Prinzip arbeiten, werden als *Modulationsverstärker* bezeichnet. In Abhängigkeit vom gewählten Modulationsverfahren unterscheidet man zwischen spannungsempfindlichen und ladungsempfindlichen Modulationsverstärkern. Im Falle der spannungsempfindlichen Modulationsverstärker spricht man von *Trägerfrequenzverstärkern*, wenn eine sinusförmige Modulationsspannung verwendet wird, und von *Zerhackerverstärkern* bei einer rechteckförmigen Modulationsspannung. Als empfindliche Gleichspannungsverstärker für meßtechnische Zwecke sind vor allem die Zerhackerverstärker wichtig. Eine mögliche Schaltungsvariante für einen solchen Verstärker ist in Abb. 7.4.1–22 dargestellt. Der Modulator und der Demodulator des Systems bestehen jeweils aus einem Schalter und einem

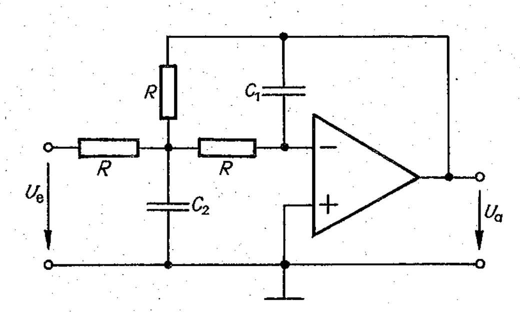


Abb. 7.4.1–21 Aktiver Tiefpaß

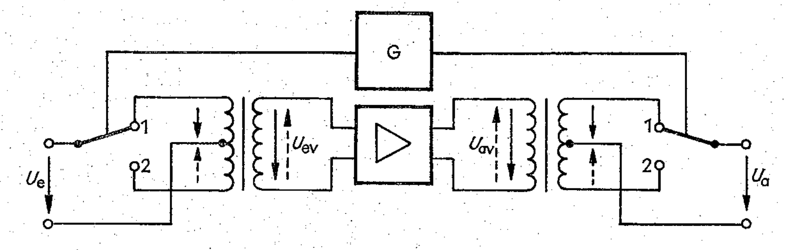


Abb. 7.4.1–22 Modulationsverstärker mit Zerhacker

Übertrager. Die beiden Schalter werden durch einen gemeinsamen Generator G betrieben und arbeiten somit synchron, d. h., beide Schalter befinden sich entweder in Stellung 1 oder Stellung 2. Die Modulatorschaltung ist so aufgebaut, daß sich bei einer Umschaltung des Schalters aus der Stellung 1 in die Stellung 2 die Stromrichtung in der Sekundärwicklung des Übertragers gerade umkehrt. Damit entsteht aus der am Eingang des Modulators liegenden Gleichspannung U_e eine Rechteckspannung U_{ev}, die dem Eingang des Verstärkers zugeführt wird und deren Frequenz durch die Schaltfrequenz des Schalters geregelt wird. Für die Verstärkung dieser Spannung werden nichtinvertierende Verstärker vom in Abb. 7.4.1–17b gezeigten Typ verwendet, die sich durch einen hohen Eingangswiderstand und wegen des vorhandenen Gegenkopplungszweiges auch durch gute Stabilität auszeichnen. Die rechteckförmige Ausgangsspannung $U_{av} = v_u U_{ev}$ wird mit Hilfe des Demodulators wieder in eine Gleichspannung U_a umgewandelt. Durch den synchronen Antrieb von Modulator und Demodulator und den Einsatz eines nichtinvertierenden Verstärkers wird erreicht, daß das Ausgangssignal U_a das Eingangssignal U_e vorzeichenrichtig wiedergibt.

Als Schalter können entweder mechanische Schalter oder Transistoren (vgl. 8.2.4.) verwendet werden. Steht noch die Forderung nach einem hohen Eingangswiderstand des Systems, so werden Feldeffekttransistoren als Schalter eingesetzt. Bei entsprechenden Maßnahmen zur Minderung von Temperatureinflüssen auf das System können Modulationsverstärker mit Zerhackern, die mit Feldeffekttransistoren bestückt sind, eine Spannungsauflösung von 0,1 µV erreichen. Derartige Verstärker dienen beispielsweise als Eingangsverstärker in hochempfindlichen Digitalvoltmetern (vgl. 7.4.1.7.).

Als ladungsempfindlicher Modulator hat der *Schwingkondensator* für die leistungslose Messung kleiner Spannungen besondere Bedeutung erlangt. Die Schaltung des Modulators ist in Abb. 7.4.1–23 gezeigt. Der als veränderlich gekennzeichnete Kondensator ist ein Plattenkondensator und hat im Ruhezustand die Kapazität $C = \varepsilon A/d$, wenn A die Plattenfläche und d der Plattenabstand ist (vgl. 8.1.1.2.). Er trägt die konstante Ladung $Q = CU_e$. Wenn man nun eine der Kondensatorplatten so zu mechanischen Schwingungen anregt, daß sich der Plattenabstand nach der Beziehung $d = d_0 + d_1 \sin \omega t$ ändert, so führt das zu einer zeitlichen Änderung seiner Kapazität gemäß $C(t) = \varepsilon A/(d_0 + d_1 \sin \omega t)$ und damit auch zu einer zeitlichen Änderung der am Kondensator liegenden Spannung um $\Delta U(t) = -(U_e/C)\Delta C(t)$. Macht man noch die in allen praktischen Schwingkondensatorausführungen erfüllte Annahme, daß $d_1 \ll d_0$ ist, so ergibt sich schließlich $\Delta U(t) = (d_1/d_0) U_e \sin \omega t$.

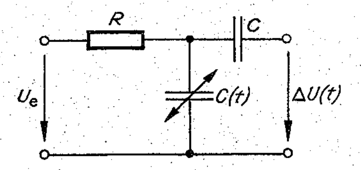


Abb. 7.4.1–23 Modulator mit Schwingkondensator

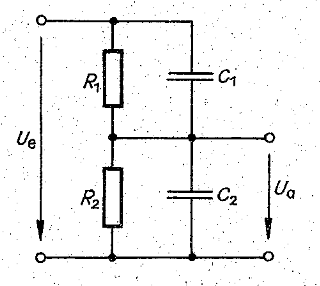


Abb. 7.4.1–24 Frequenzunabhängiger Spannungsteiler

Diese Wechselspannung wird verstärkt und anschließend wieder demoduliert oder einfach gleichgerichtet. Mit Schwingkondensatoren erreicht man Eingangswiderstände bis zu $\approx 10^{15}\,\Omega$ und kann noch Spannungen im Bereich von 10^{-9} V messen. Strommessungen sind bis herunter zu 10^{-16} A möglich. Bei nicht zu hohen Anforderungen an die Höhe des Eingangswiderstands des Modulationsverstärkers können auch Kapazitätsdioden (vgl. 8.1.3.2.) die Aufgabe des Schwingkondensators übernehmen.

Die Eingangsschaltung der Spannungsmeßgeräte mit Verstärkern hängt von der Anwendung des Meßgerätes ab und kann einen Schutz gegen zu hohe Eingangsspannungen, eine Verstärkerstufe zur Erzeugung eines hohen Eingangswiderstandes (etwa eine Elektromotorschaltung nach Abb. 7.4.1–17) oder eine Gleichrichterschaltung enthalten, die eine anliegende Meßwechselspannung vor der Verstärkung gleichrichtet. Die Gleichrichtung einer zu messenden Wechselspannung ist immer dann erforderlich, wenn die Frequenz der Wechselspannung über der oberen Grenzfrequenz des verwendeten Verstärkers liegt. Der der Eingangsschaltung folgende Spannungsteiler hat die Aufgabe, die zu messende Spannung dem Verstärker anzupassen, d. h. den Spannungspegel am Verstärkereingang so einzustellen, daß der Verstärker im linearen Teil seiner Übertragungskennlinie angesteuert wird. Dieser Spannungsteiler muß frequenzunabhängig sein. Eine mögliche Schaltung für einen solchen Spannungsteiler zeigt Abb. 7.4.1–24. Bei Einhaltung der Bedingung $R_1C_1 = R_2C_2$ ergibt sich für den Frequenzgang dieser Schaltung $G(i\omega) = U_a/U_e = R_2/(R_1 + R_2)$.

Zur Messung von Wechselspannungen werden Breitbandverstärker oder selektive Verstärker eingesetzt, im Falle von Gleichspannungen können auch Modulationsverstärker verwendet werden. In analog arbeitenden Meßgeräten werden zur Anzeige des Meßwertes vorwiegend gleichspannungsanzeigende Meßinstrumente verwendet, also etwa Drehspulinstrumente (vgl. 7.4.1.1.) oder registrierende Meßgeräte (vgl. 7.4.1.5.). Verstärkte Wechselspannungen müssen folglich vorher gleichgerichtet werden, etwa mit der in Abb. 7.4.1–3 angegebenen Schaltung.

7.4.1.5. Registrierende Meßgeräte dienen der Aufzeichnung einer Meßgröße als Funktion der Zeit oder als Funktion einer zweiten Meßgröße. Die hierfür üblicherweise verwendeten Geräte sind der *Motorkompensator* (*Kompensationsbandschreiber*) und der *Koordinatenschreiber* (*x-y-Schreiber*).

Motorkompensator. Abb. 7.4.1–25 zeigt die Prinzipschaltung eines Motorkompensators für den Fall einer Spannungsmessung. Der wesentliche Teil eines Motorkompensators ist ein selbstabgleichender Poggendorff-Kompensator, von dem hier nur die gesuchte Spannung U_x und der zum Abgleich dienende Widerstand R dargestellt sind (vgl. Abb. 7.4.1–9); alle für die Erzeugung und Einstellung des Hilfsstromes erforderlichen Schaltelemente sind in Abb. 7.4.1–25 weggelassen. Der Widerstand R ist hier ein kontinuierlich einstellbares Meßpotentiometer. Die zu messende Spannung U_x wird mit der Spannung U_v verglichen, die in Abhängigkeit von der gerade vorhandenen Einstellung des Meßpotentiometers i. allg. einen von U_x verschiedenen

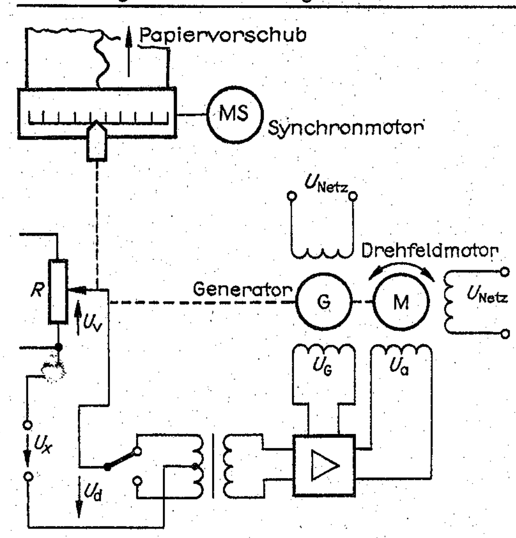


Abb. 7.4.1-25 Motorkompensator

Wert haben wird. Zur Erzielung des Abgleichs ist die Differenzspannung $U_d = U_v - U_x$ durch entprechende Verstellung des Potentiometers zu Null zu machen. Als Nullindikator (Nullinstrument G in Abb. 7.4.1-9) dient ein Verstärker, dem die Differenzspannung U_d zugeführt wird. Im Verstärkereingang liegt ein mit Netzfrequenz betriebener Zerhakker (meist mit mechanischem Schalter), der aus der Gleichspannung U_d eine Rechteckspannung mit einer Frequenz von 50 Hz erzeugt. Der Zerhacker ist dabei bezüglich seiner Phasenlage so eingestellt, daß die Ausgangsspannung U_a des Verstärkers, die ebenfalls eine Rechteckspannung mit einer Frequenz von 50 Hz ist, bei positiver Eingangsspannung U_d gerade eine Phasenverschiebung von $\varphi = 90°$ zur Netzspannung U_{Netz} zeigt. Eine negative Eingangsspannung U_d führt demzufolge zu einer negativen Phasenverschiebung von $\varphi = -90°$. Nun wird die Spannung U_a der einen, die Netzspannung U_{Netz} der anderen Wicklung eines Drehfeldmotors M zugeführt. Das auf den Rotor des Drehfeldmotors ausgeübte Drehmoment M_r ist dann $M_r \sim |U_{Netz}|\,|U_a| \sin \varphi$. Folglich führt der durch eine Vorzeichenänderung von U_d bedingte Vorzeichenwechsel der Phasenverschiebung φ zu einer Umkehr der Drehrichtung des Rotors des Drehfeldmotors. Koppelt man die Achse des Drehfeldmotors mit der Schleiferachse des Meßpotentiometers, so läßt sich dadurch U_v in der gewünschten Richtung verändern, da die Einstellung des Potentiometers stets in Richtung $U_d \to 0$ erfolgt. Wird $U_d = U_v - U_x = 0$ und damit auch $U_a = 0$, so ist der Abgleich des Kompensators erreicht und der Drehfeldmotor bleibt stehen.

Zur Registrierung des Meßsignals koppelt man die Achse des Meßpotentiometers mit einem Zeiger, an dem eine Schreibfeder befestigt ist. Mit Hilfe des Zeigers kann man einmal auf einer entsprechend geeichten Skale die Größe von U_x ablesen, zum anderen wird das Registrierpapier durch einen Synchronmotor MS mit Getriebe unter der Schreibfeder vorbeibewegt, so daß der Vorschub des Papiers die Zeitachse bildet und somit $U_x(t)$ aufgezeichnet wird. Über das Übersetzungsverhältnis des Getriebes kann der Maßstab der Zeitachse variiert werden. Zur Erzielung einer optimalen Einstellgeschwindigkeit des Systems wird der Einstellmechanismus gedämpft. Das geschieht dadurch, daß man die Rotorachse des Drehfeldmotors M starr mit der Rotorachse eines Generators G koppelt. Der Generator erzeugt eine der Drehgeschwindigkeit seines Rotors proportionale Spannung U_G, die dem Verstärker als Gegenkopplungsspannung zugeführt wird. Hiermit läßt sich die Dämpfung des Einstellvorgangs optimal gestalten. Wegen seiner großen Genauigkeit (Kompensationsverfahren), großen Einstellsicherheit und des geringen Bedarfes an Meßenergie ist der Motorkompensator ein universell einsetzbares Meßgerät. Es können alle physikalischen Größen gemessen werden, die sich als Spannungs-, Strom- oder Widerstandsänderung darstellen lassen.

Sind mehrere sich langsam ändernde Meßgrößen parallel aufzuzeichnen, so kann anstelle des beschriebenen *Linienschreibers* ein *Punktschreiber* eingesetzt werden. Er unterscheidet sich vom Linienschreiber dadurch, daß die Schreibfeder durch einen Druckkopf mit mehreren Schreibfedern ersetzt wird, die jedes Signal in einer anderen Farbe aufzeichnen. Gleichzeitig mit dem Umschalten des Eingangs U_x auf eine andere Meßstelle erfolgt auch ein Umschalten auf die entsprechende Schreibfeder. Es gibt praktische Ausführungsformen von Punktschreibern, mit denen es möglich ist, bis zu zwölf Meßgrößen gleichzeitig aufzuzeichnen.

Der *Koordinatenschreiber* hat zwei sowohl elektrisch als auch mechanisch vollkommen voneinander unabhängig arbeitende Systeme vom in Abb. 7.4.1-25 dargestellten Typ, so daß zu jedem Zeitpunkt zwei Spannungen gleichzeitig gemessen werden können. Mit Hilfe eines geeignet konstruierten Schreiberschlittens kann die eine der gemessenen Spannungen als Auslenkung in x-Richtung, die andere als Auslenkung in y-Richtung auf einem festgehaltenen Registrierpapier aufgezeichnet werden. Auf diese Weise läßt sich der funktionelle Zusammenhang beider Spannungen ermitteln.

7.4.1.6. Elektronenstrahloszillograf. Die Hauptbaugruppen eines Oszillografen sind die Oszillografenröhre (vgl. 8.1.2.3.), der x- und y-Verstärker, der Zeitablenkgenerator, die Triggereinheit sowie das Netzteil, welches alle erforderlichen Versorgungsspannungen liefert. Da die zu messenden Signale sowohl bezüglich ihrer Amplituden als auch ihres Frequenzspektrums sehr unterschiedlich sein können, werden an die verwendeten Verstärker sehr hohe Anforderungen bezüglich ihres statischen (keine Nullpunktdrift) und dynamischen Verhaltens (konstanter Amplituden- und Phasengang über den gesamten Frequenzbereich) gestellt. Es können demzufolge nur Breitbandverstärker (vgl. 7.4.1.4.) hoher Güte eingesetzt werden.

Häufig will man den zeitlichen Verlauf einer Spannung, etwa der dem y-Eingang zugeführten Spannung $U_y(t)$, ermitteln. Dann wird am x-Eingang eine zeitlineare Spannung $U_x(t) = at$ benötigt. Sie wird von einem Ablenkgenerator in Form eines Sägezahnimpulses geliefert. Während der Anstiegszeit des Sägezahnimpulses erfolgt die zeitliche Darstellung der Meßgröße $U_y(t)$. In der darauf folgenden kurzen Abfallzeit von U_x wird der Elektronenstrahl in seine Anfangslage zurückgeführt. Damit der Rücklauf des Elektronenstrahles das Oszillografenbild nicht stört, wird der Elektronenstrahl während dieser Zeit dunkelgetastet. Danach steigt U_x wieder an und $U_y(t)$ wird erneut auf dem Leuchtschirm dargestellt. Dieser Vorgang wird periodisch wiederholt. Damit jedoch auf dem Leuchtschirm ein stehendes Bild entsteht, muß der Schreibvorgang nach jedem Rücklauf des Elektronenstrahles wieder mit dem gleichen

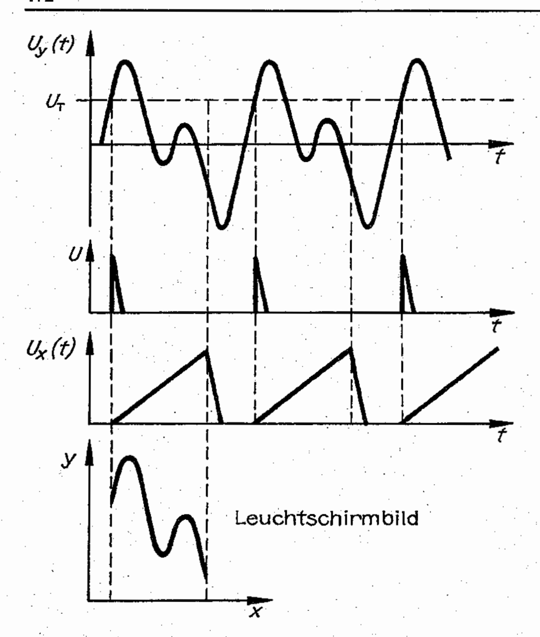
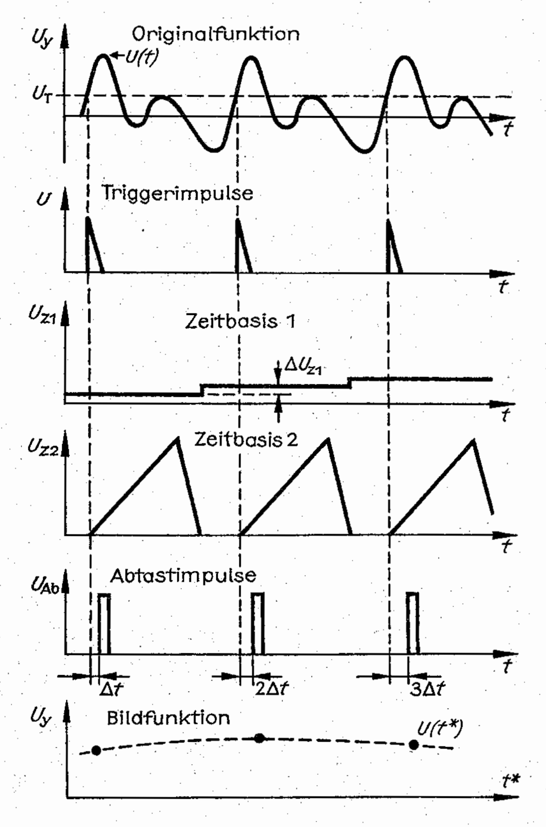


Abb. 7.4.1–26 Spannungs-Zeit-Diagramm eines Oszillografen mit Trigger



Momentanwert der zu messenden periodischen Spannung $U_y(t)$ beginnen. Das bedeutet, daß zwischen der Meßspannung $U_y(t)$ und der Zeitablenkspannung $U_x(t)$ frequenz- und phasenmäßige Übereinstimmung bestehen muß. Dieser Gleichtakt zwischen U_y und U_x wird mit Hilfe eines *Triggersystems* erreicht. Das Prinzip der Gleichtaktherstellung veranschaulicht Abb. 7.4.1–26. Durchläuft die Meßspannung $U_y(t)$ einen am Oszillografen einstellbaren Triggerpegel U_T, so wird durch eine Triggerschaltung ein Spannungsimpuls U abgegeben, der die Auslösung des Ablenkgenerators bewirkt. Während der Anstiegszeit des Sägezahnimpulses $U_x(t)$ wird die Meßspannung auf dem Leuchtschirm dargestellt. Nach dem Rücklauf des Elektronenstrahles in seine Ausgangsstellung erfolgt erst dann eine erneute Auslösung des Ablenkgenerators, wenn der nächste Triggerimpuls auftritt. Damit wird erreicht, daß stets derselbe Ausschnitt des Meßsignals auf dem Leuchtschirm erscheint, es entsteht ein stehendes Bild. Durch unterschiedliche Wahl des Triggerpegels und der Anstiegszeit des Sägezahnimpulses kann man jeden beliebigen Ausschnitt des Meßsignals auf dem Leuchtschirm abbilden.

Mit dem Oszillografen können auch einmalige oder nichtperiodische Vorgänge untersucht werden. Außerdem kann man mit einem Oszillografen ähnlich wie mit einem Koordinatenschreiber (vgl. 7.4.1.5.) auch den funktionellen Zusammenhang zweier Meßspannungen untersuchen. Dazu werden die Meßspannungen dem x- und y-Eingang des Oszillografen zugeführt.

Die obere Grenzfrequenz des beschriebenen sogenannten *Echtzeitoszillografen* läßt sich jedoch nicht beliebig erhöhen. Die Ursache hierfür liegt einerseits in der endlichen Laufzeit der Elektronen durch die Ablenkplatten (vgl. 8.1.2.3.), zum anderen darin, daß das Produkt aus Bandbreite und Verstärkung eines Systems eine Konstante ist. Bei hinreichend hoher Eingangsempfindlichkeit des verwendeten Meßverstärkers ist eine obere Grenzfrequenz von $f_{g0} \approx 500$ MHz technisch noch realisierbar. Für die Messung von Signalen, die höhere Frequenzanteile enthalten, wird der sogenannte *Abtastoszillograf* (*Sampling-Oszillograf*) verwendet. Die Anwendbarkeit des Abtastoszillografen ist jedoch auf periodische Signale oder auf Signale mit identischem Kurvenverlauf, sofern sie zu unterschiedlichen Zeitpunkten auftreten, beschränkt. Zufällige oder einmalig auftretende Signale können mit dem Abtastoszillografen nicht gemessen werden. Das Grundprinzip des Abtastoszillografen besteht darin, daß innerhalb einer Periode des Meßsignals maximal ein Momentanwert des Meßsignals abgetastet und als Bildpunkt auf dem Leuchtschirm dargestellt wird. Wählt man nun während einer hinreichend großen Zahl von Perioden des Meßsignals die Abtastzeitpunkte so, daß sie bei zwei aufeinander folgenden Abtastvorgängen jeweils um eine bestimmte Zeitspanne Δt gegeneinander zeitlich verschoben sind, so entsteht auf dem Leuchtschirm eine Folge von Meßpunkten, die den ursprünglichen Verlauf des Meßsignals wiedergibt, aber eine verlängerte Darstellzeit hat. Es erfolgt eine Zeittransformation, die durch die Zahl der Abtastimpulse, d. h., die Zahl der abgetasteten Perioden des Meßsignals festgelegt ist.

Die praktische Realisierung des Abtastoszillografen veranschaulicht Abb. 7.4.1–27. Das Meßsignal $U_y(t)$ wird auf den y-Eingang des Oszillografen gegeben. Durchläuft nun das

Abb. 7.4.1–27 Spannungs-Zeit-Diagramm des Abtastoszillografen

Meßsignal den einstellbaren Triggerpegel U_T, so liefert wie beim Echtzeitoszillografen (vgl. Abb. 7.4.1–26) eine Triggerschaltung einen Triggerimpuls U, der den Ablenkgenerator (die sogenannte Zeitbasis 2) auslöst, der die Sägezahnspannung U_{z2} erzeugt. Diese Spannung U_{z2} wird nun mit der durch einen Treppengenerator (die sogenannte Zeitbasis 1) erzeugten Spannung U_{z1} verglichen. Wenn $U_{z2} = U_{z1}$ wird, was nach einer Zeitspanne Δt nach Auslösung des Sägezahnimpulses geschieht, wird der Elektronenstrahl durch einen Abtastimpuls für eine sehr kurze Zeit (die Abtastzeit muß wesentlich kürzer als die Periode des Meßsignals sein) hellgetastet. Es entsteht ein Bildpunkt auf dem Leuchtschirm, der einen einzigen Momentanwert des Meßsignals wiedergibt. Nach beendetem Anstieg der Sägezahnspannung U_{z2} wird die Spannung U_{z1} durch den Treppenspannungsgenerator um einen konstanten Spannungsbetrag ΔU_{z1} erhöht. Nach dem Eintreffen des nächsten Triggerimpulses, der bezüglich des Meßsignals die gleiche Phasenlage hat wie der vorhergehende Triggerimpuls, wird der die Sägezahnspannung U_{z2} liefernde Ablenkgenerator erneut ausgelöst. Die für die Auslösung des Abtastimpulses erforderliche Bedingung $U_{z2} = U_{z1}$ wird jetzt jedoch erst zu einem Zeitpunkt erreicht, der um die Zeitspanne $2\Delta t$ gegenüber dem Triggersignal verschoben ist. Damit wird ein Momentanwert des Meßsignals auf dem Leuchtschirm abgebildet, der im Meßsignal um die Zeitspanne Δt später auftritt als der vorhergehende abgebildete Momentanwert. Auf dem Leuchtschirm erscheint dieser Momentanwert jedoch um die Zeitspanne $T_0 + \Delta t$ später als der vorhergehende Momentanwert, wenn T_0 die Periode des Meßsignals ist. Der Abtastvorgang wird nun solange wiederholt, bis die Spannung U_{z1} den Maximalwert der Sägezahnspannung U_{z2} erreicht hat. Damit bestimmt die Dauer des Sägezahnimpulses die Größe des dargestellten Ausschnittes des Meßsignals, während durch die Wahl der Stufenhöhe ΔU_{z1} die Zahl der Abtastimpulse und damit die Zahl der Punkte der Leuchtschirmkurve festgelegt wird. In Abb. 7.4.1–27 sind drei aufeinander folgende Abtastvorgänge dargestellt, wodurch die Bildfunktion gerade den rot eingezeichneten Teil $U(t)$ der Originalfunktion, d. h., des zu messenden Signals wiedergibt.

Im Abtastoszillografen wird eine kontinuierliche Funktion in eine diskontinuierliche Meßwertfolge umgewandelt.

Soll diese Umwandlung ohne Informationsverlust erfolgen, so ist hierfür eine bestimmte Mindestzahl von Abtastpunkten erforderlich, d. h., das Zeitintervall Δt muß unterhalb eines bestimmten Wertes liegen. Aus dem Abtasttheorem (vgl. 7.4.1.4.) folgt, daß zur Darstellung einer Funktion, deren Spektrum eine höchste Frequenz f_{g0} aufweist, die Abtastwerte in einem zeitlichen Abstand $\Delta t \leqq 1/2f_{g0}$ gewählt werden müssen.

Eine Erweiterung des Abtastoszillografen stellt das *Boxcar-Verfahren* (Boxcar-Integrator) dar. Während beim Abtastoszillografen jeder Momentanwert des Meßsignals nur einmal abgetastet wird, wird beim Boxcar-Verfahren jeder Meßwert N-mal abgetastet und der lineare Mittelwert dieser Abtastungen gebildet. Daraus ergibt sich eine Verbesserung des Signal-Rausch-Verhältnisses um den Faktor $\sqrt{N}$, so daß dieses Verfahren sich besonders für die Messung stark verrauschter Signale eignet.

Die gleichzeitige Darstellung von zwei Meßsignalen $U_{y1}(t)$ und $U_{y2}(t)$ ist mit Zweistrahl- oder Zweikanaloszillografen möglich. Beim *Zweistrahloszillografen* hat die Oszillografenröhre zwei Strahlsysteme und jeweils zwei x- und y-Ablenkplattenpaare, so daß die beiden Meßsignale tatsächlich unabhängig voneinander (auch mit unterschiedlichen Zeitmaßstäben) dargestellt werden können. Der *Zweikanaloszillograf* hat eine Einstrahlröhre. Die beiden Meßsignale werden durch zwei unterschiedliche Verstärker zunächst separat verstärkt und dann mit Hilfe eines elektronischen Schalters abwechselnd dem y-Eingang der Oszillografenröhre zugeführt.

Man unterscheidet zwei Betriebsarten des Zweikanaloszillografen. Bei der Betriebsart „chopper" ist die Umschaltzeit des elektronischen Schalters wesentlich kürzer als die Anstiegszeit des Sägezahnimpulses, so daß beide Meßsignale als unterbrochene Kurven gleichzeitig auf dem Bildschirm erscheinen. Bei hinreichend hoher Umschaltfrequenz (≈ 500 kHz) erscheinen beide Signale aufgrund der Trägheit des menschlichen Auges und der endlichen Auflösung des Leuchtschirms als geschlossene Kurvenzüge. Bei sehr kurzen Ablenkzeiten werden die Lücken in den Kurven sichtbar und können sich auch bei der Auswertung nachteilig auswirken. In solchen Fällen wählt man die Betriebsart „alternierend", bei der die Umschaltung jeweils nach Beendigung einer Periode des Ablenksignals erfolgt. Die beiden durch die Meßsignale hervorgerufenen Kurvenzüge erscheinen folglich nacheinander auf dem Bildschirm. Bei hohen Ablenkfrequenzen führt die Trägheit des Auges dazu, daß beide Kurvenzüge praktisch gleichzeitig gesehen werden.

Moderne Oszillografen sind vielfach mit elektronischen Speichern ausgestattet. Man spricht dann von *Speicheroszillografen*. Sie bieten den Vorteil, daß die einmal gespeicherten Meßsignale auch zu einem späteren Zeitpunkt wieder reproduziert werden können. Außerdem können sie jederzeit elektronisch weiterverarbeitet werden.

7.4.1.7. Digitale Meßgeräte. Neben den Impulszählern (vgl. 8.2.4.4.), die auch für die Frequenz- und Zeitmessung eingesetzt werden können, ist das *Digitalvoltmeter* das wichtigste digitale Meßgerät. Es dient der Messung von Gleichspannungen. Das Eingangssignal durchläuft zunächst ein Tiefpaßfilter, welches äußere Störspannungen (Wechsel-, Impuls- und Rauschspannungen) unterdrückt. Ein Eingangsspannungsteiler hat die Aufgabe, das Eingangssignal an den Aussteuerbereich des nachfolgenden Verstärkers anzupassen (vgl. 7.4.1.4.). Häufig erfolgt die Umschaltung des Eingangsteilers selbsttätig durch eine Steuereinheit, d. h., das Meßgerät wählt sich den richtigen Meßbereich selbst automatisch aus. Die Wahl des Verstärkers hängt von den Genauigkeitsansprüchen an das Digitalvoltmeter ab. Für hochempfindliche Geräte werden Zerhackerverstärker mit Feldeffekttransistoren (vgl. 7.4.1.4.) verwendet. Die wichtigste Funktionseinheit des Digitalvoltmeters ist der *Analog-Digital-Umsetzer*, denn seine Parameter bestimmen ganz wesentlich die Meßeigenschaften des Digitalvoltmeters. Die indirekten Umsetzverfahren basieren auf der Zwischenumsetzung der zu messenden Spannung in eine Zeit, die dann mit Hilfe eines Impulszählers bestimmt wird (vgl. 8.2.5.). Bei höheren Genauigkeitsansprüchen werden direkte Umsetzverfahren verwendet, bei denen die verstärkte Eingangsspannung unmittelbar, ohne Zwischenschaltung einer anderen Größe, mit einer Vergleichsspannung verglichen wird. Diese sogenannten *Kompensationsumsetzer* arbeiten also nach dem gleichen Prinzip wie ein Gleichspannungskompensator (vgl. 7.4.1.2.). Die Kompensationsspannung U_k wird von einem Digital-Analog-Wandler geliefert, der aus einer Referenzspannungsquelle U_N (etwa einem Weston-Element, vgl. 7.4.1.2.) und einem Spannungsteiler besteht, der beispielsweise so aufgebaut sein kann, daß die mit Hilfe elektronischer Schalter einstellbaren Vergleichsspannungen

in Stufen bereitgestellt werden, die dem BCD-Code (vgl. 8.2.4.3.) entsprechen. Der Vergleich zwischen der verstärkten Eingangsspannung U_x des Digitalvoltmeters und der Vergleichsspannung U_k erfolgt in einem als Komparator geschalteten Differenzverstärker (vgl. 8.2.5.). Die Ausgangsspannung des Differenzverstärkers wird der Steuerschaltung zugeführt, die den Spannungsteiler solange steuert, bis die Differenz zwischen den beiden Spannungen kleiner als die halbe Quantisierungseinheit des Umsetzers ist. Das Ergebnis des Abgleiches kann entweder über einen Informationsausgang in codierter Form einer elektronischen Weiterverarbeitung oder nach einer Decodierung (vgl. 8.2.4.3.) einer Ziffernanzeige zugeführt werden.

Werden der Eingangsschaltung des Digitalvoltmeters entsprechende Wandler vorgeschaltet, so können Digitalvoltmeter auch zur Messung von Gleichströmen, Wechselspannungen, Wechselströmen und Widerständen eingesetzt werden. Geräte, die mit entsprechenden zusätzlichen Eingangsstufen bestückt sind, werden als *Digitalmultimeter* bezeichnet. Für die Gleichstrommessung wird ein Strom-Spannungs-Wandler vorgeschaltet, wie er in Abb. 7.4.1–18 dargestellt ist. Bei der Messung von Wechselspannungen

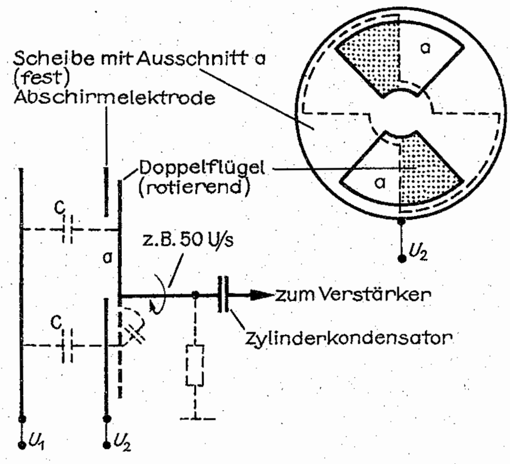


Abb. 7.4.1–28 Rotationsvoltmeter, Anordnung meist im Druckkessel (N_2, Edelgase o. ä.), um Durchschläge bei hohen Spannungen zu verhindern

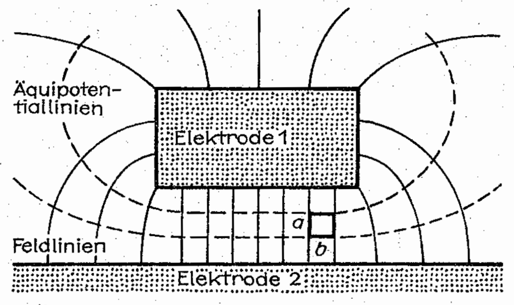


Abb. 7.4.1–29 Grafische Konstruktion eines ebenen elektrischen Feldes

oder Wechselströmen ist eine Gleichrichtung erforderlich. Dafür kann die in Abb. 7.4.1–3 gezeigte Schaltung verwendet werden, die zur Bildung des Gleichrichtwertes des Wechselsignals führt. Die Größe Ohmscher Widerstände wird durch die Messung des Spannungsabfalls bestimmt, die ein bekannter Strom aus einer Konstantstromquelle an diesem Widerstand erzeugt.

7.4.1.8. Messung elektrischer Felder. Zur experimentellen Bestimmung von elektrischen Feldstärken kann man u. a. sogenannte *Rotationsvoltmeter* verwenden, die eine Spannungsdifferenz $U_1 - U_2$ mittels einer Abschirm- und einer rotierenden Influenzelektrode (kleiner Motor oder Tauchspule, ≈ 50 Umdrehungen/s) und einem Innenwiderstand $R_i \gtrsim 10^{12}\,\Omega$ messen. Die Wirkungsweise beruht darauf, daß die Meßelektrode periodisch vom wirkenden Feld abgeschirmt wird und die influenzierte Ladung der Feldstärke proportional ist. Die Influenzladung mißt man mit spannungsempfindlichen Verstärkern. Eine solche Anordnung ist in Abb. 7.4.1–28 skizziert. Es lassen sich Feldstärken zwischen ≈ 100 V/m bis zu vielen 10^5 V/m messen. Die Kalibrierung erfolgt durch ein Parallelfeld bekannter Feldstärke.

Elektrische Wechselfelder werden weitestgehend über Spannungsmessungen bestimmt.

In Wechselfeldern läßt sich die elektrische Feldstärke auch aus der thermisch gut meßbaren Strahlungsleistung W je Fläche ermitteln. Im Vakuum (in guter Näherung auch in Luft) gilt $E = \sqrt{2W/(\varepsilon_0 c)}$. Für die die Erde außerhalb der Lufthülle erreichende Sonnenstrahlung erhält man $E = 1010$ V/m (für die magnetische Feldstärke folgt aus der analogen Gleichung $H = \sqrt{2W/(\mu_0 c)} = 2{,}68$ A/m.

Die Potentialflächen in elektrischen Feldern kann man näherungsweise auch durch grafische Methoden erhalten. Als Beispiel diene eine ebene Feldverteilung. Zeichnet man die Feld- und Äquipotentiallinien so, daß der mittlere Linienabstand an jeder Feldstelle bei beiden Kurvenscharen gleich ist, so entsteht eine orthogonale Netzstruktur, da sich die Scharen rechtwinklig schneiden (Abb. 7.4.1–29). Man zeichnet den Verlauf der genannten Linien und verbessert so lange, bis die orthogonale Struktur erreicht ist. In jedem Viereck ist dann die mittlere Breite gleich der mittleren Länge, wobei sich die Kurven stets rechtwinklig schneiden müssen.

Zur experimentellen Bestimmung der Potentialverteilung einer Anordnung von Elektroden verwendet man vorteilhaft den *elektrolytischen Trog*. Da sich Potential- und Feldverteilung nicht ändern, wenn proportionale Maßstabsänderungen der Anordnung erfolgen, kann man auch Modelle (vergrößerte oder verkleinerte) untersuchen. Der elektrolytische Trog besteht aus einer Schale aus isolierendem Material, die mit einem schwachen Elektrolyten so hoch gefüllt ist, daß die Elektroden genügend bedeckt sind. Die Messung der Äquipotentialflächen erfolgt mit stromloser Sonde S, d. h. $I = 0$. Der Abgleich erfolgt mit dem Potentiometer P und dem Detektor D (Abb. 7.4.1–30). Das Potential an den Oberflächen der Elektroden ist bekannt. Eine bestimmte Äquipotentialfläche gewinnt man, indem die Sonde so durch den Trog geführt wird, daß für die eingestellte Spannung am Potentiometer der Sondenstrom stets null bleibt. Diese Sondenbewegung wird mit Hilfe der Schlittenführung auf einen Zeichenstift übertragen. Prinzipiell sind auch räumliche Messungen möglich. Zur Vermeidung von Kontaktpotentialen an der Sonde und an den Elektroden betreibt man den Trog meist mit niederfrequentem Wech-

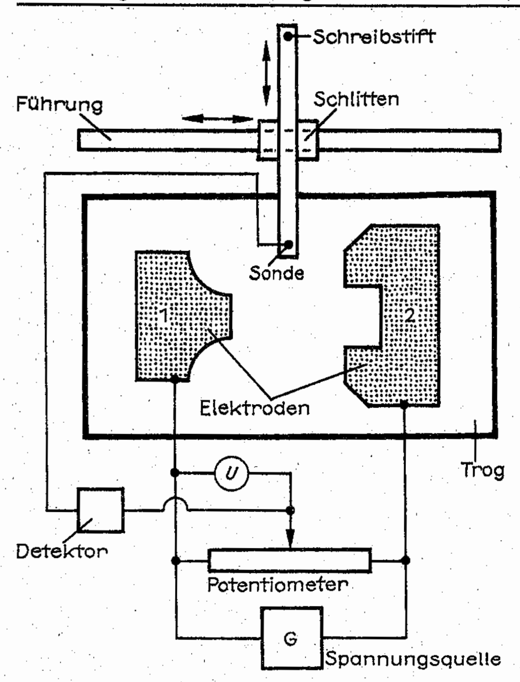


Abb. 7.4.1–30 Elektrolytischer Trog

selstrom. Auf den ermittelten Äquipotentialflächen müssen die Feldlinien senkrecht stehen. Sie werden in die erhaltene Kurvenschar eingezeichnet.
Analog erhält man mittels fester oder rotierender Sonden z. B. die Potentialverteilung in Gasentladungen.
Eine weitere Möglichkeit zur Ermittlung der Potentialverhältnisse insbesondere in Plasmaentladungen ergibt sich aus der Reflexion und Transmission von Mikrowellen.

7.4.2. Messung magnetischer Felder

Anordnungen zur Messung von magnetischen Feldern oder zur Bestimmung der magnetischen Eigenschaften der Materie heißen allgemein *Magnetometer*.

Stromwaage. Eine absolute Bestimmung von B (d. h. unmittelbarer Anschluß an andere Basiseinheiten) ist mit der in Abb. 7.4.2–1 gezeigten Stromwaage möglich. Dort wird die Kraft F eines stromdurchflossenen Leiters im Magnetfeld der Induktion B genutzt. Am Waagebalken W befindet sich das Leiterstück der Länge l, das vom Strom I durchflossen ist. Die Zuleitung des elektrischen Stromes erfolgt über Quecksilberkontakte am Drehpunkt des Hebelarms (hier nicht gezeichnet). Es gilt $F = B \cdot l \cdot I$, $(I \perp B)$; für $F = 10^{-2}$ N, $l = 0{,}02$ m, $I = 1$ A erhält man $B = 0{,}5$ T; l muß innerhalb des homogenen Feldes bleiben.
Stromwaagen sind relativ kompliziert und dienen i. allg. Kalibrierungs- bzw. Eichzwecken. Je nach Ausführung der Apparatur und des Meßbereiches werden Genauigkeiten $\Delta B/B$ bis $\approx 10^{-4}$ erreicht (in Einzelfällen auch noch kleinere Werte).

Messung mit Hilfe eines Stabmagneten. Eine einfache Anordnung zur Bestimmung der Stärke eines magnetischen Feldes mit Hilfe eines Stabmagneten ist in Abb. 7.4.2–2 skizziert. Dazu führt man folgende Messungen aus:

a) Der Stabmagnet wird freibeweglich in dem zu bestimmenden Feld $\boldsymbol{H}$ (homogen vorausgesetzt) möglichst torsionsfrei aufgehängt. Auf diesen Magnetstab wirkt infolge seines magnetischen Momentes $\boldsymbol{m}$ im Feld $\boldsymbol{H}$ ein Drehmoment $\boldsymbol{D} = \mu_0 \boldsymbol{m} \times \boldsymbol{H}$, so daß der Magnet um seine in Richtung von $\boldsymbol{H}$ liegende Ruhelage bei kleinen Auslenkungen harmonische Schwingungen mit der Dauer $T = 2\pi\sqrt{\Theta(\mu_0 mH)^{-1}}$ ausführt, wobei Θ das Trägheitsmoment des Stabmagneten um seine Drehachse ist. Man erhält daraus $mH = 4\pi^2\Theta/\mu_0 T^2$.
b) Das vom Stabmagneten erzeugte magnetische Feld $\boldsymbol{H}_{St}$ wird mit dem zu messenden Feld $\boldsymbol{H}$ überlagert und die Richtung des resultierenden Feldes mittels einer Magnetnadel (im einfachsten Fall: Kompaß) gemessen. Stehen nun nach Abb. 7.4.2–2 die beiden Felder $\boldsymbol{H}_{St}$ und $\boldsymbol{H}$

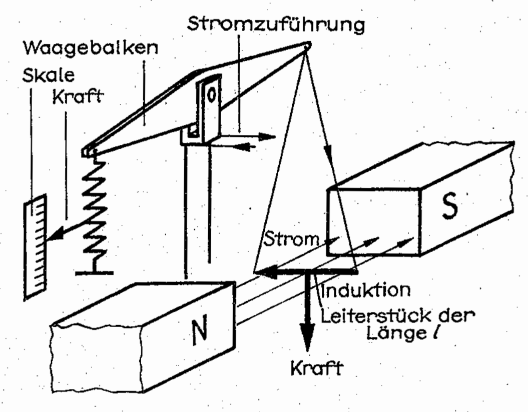


Abb. 7.4.2–1 Prinzip einer Stromwaage

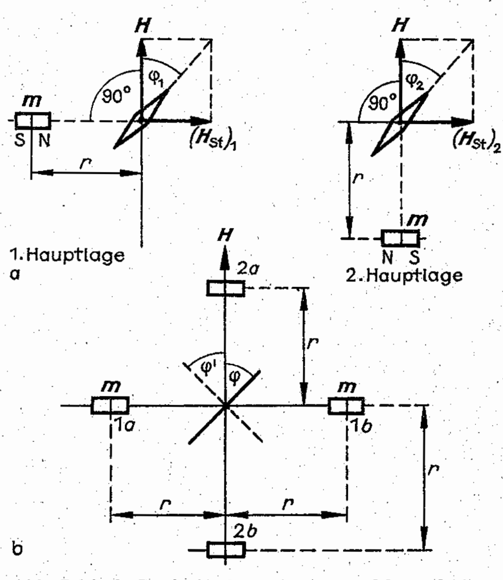


Abb. 7.4.2–2 Gaußsche Hauptlagen zur Magnetfeldmessung

senkrecht aufeinander und ist der Abstand r groß gegen die Länge des Stabmagneten und der Magnetnadel, so erhält man $m/H = 2\pi r^3 \tan\varphi_1$ (1. Hauptlage) und $m/H = 4\pi r^3 \tan\varphi_2$ (2. Hauptlage).
Kombiniert man beide Messungen a) und b) so folgen mit $m/H = Q$ und $mH = P$ die gesuchten Größen $H = \sqrt{P/Q}$ und $m = \sqrt{PQ}$. Führt man die Messungen im Erdfeld aus, so erhält man für die Horizontalintensität des erdmagnetischen Feldes in Mitteleuropa ≈ 16 A/m. Legt man nach Abb. 7.4.2–2b den Stabmagneten noch in die Lagen 1a, 1b, 2a und 2b und kehrt ihn in jeder Lage nochmals um je 180° um, so erhält man für den Winkel φ acht Werte, aus denen ein Mittelwert gebildet wird. Die Meßgenauigkeit kann noch weiter durch Korrekturen bezüglich der Polabstände beider Magneten gesteigert werden.
Tangentenbussole. Sie besteht aus einer kreisförmigen flachen Spule mit dem Radius r und n Windungen, in deren Mittelpunkt eine Magnetnadel frei schwingend angeordnet ist. Die Magnetnadel stellt sich in Richtung des zu messenden magnetischen Feldes $\boldsymbol{H}$ ein. Das magnetische Feld $\boldsymbol{H}_{Sp}$ im Mittelpunkt der Spule ergibt sich zu $\boldsymbol{H}_{Sp} = In/(2r)$ (I = Strom durch die Spule).
Stehen beide Felder $\boldsymbol{H}$ und $\boldsymbol{H}_{Sp}$ nach Abb. 7.4.2–3 senkrecht aufeinander und bezeichnet φ den Ablenkwinkel der Magnetnadel gegen die Richtung $\boldsymbol{H}$ (also des zu messenden Feldes), so ist $H = In/(2r \tan\varphi)$. Regelt man I auf $\varphi = 45°$ ein, so sind beide Felder dem Betrag nach gleich.
Magnetometer mit rotierender Spule. Wird ein Leiter L im Magnetfeld $\boldsymbol{B}$ bewegt, so entsteht nach dem Induktionsgesetz eine elektrische Spannung. Man läßt eine Meßspule auf der Achse eines schnellaufenden Motors in dem zu messenden Feld rotieren. Die induzierte Wechselspannung wird direkt oder phasengleichgerichtet angezeigt. Es sind viele Ausführungen möglich, z. B. Spulen von 2 mm Durchmesser und ≈ 500 Windungen. Die Spulengröße bestimmt den kleinsten noch zugänglichen Raum bzw. Polabstand, in dem das Magnetfeld gemessen werden kann. Für höhere Ansprüche haben solche Instrumente noch einen kleinen Permanentmagneten auf der Motorachse, der in Verbindung mit einer festen Probenspule eine Vergleichsspannung liefert, die mit der Meßspannung verglichen wird (Eliminierung von Unregelmäßigkeiten des Motors und des Anzeigekanals). Je nach Selektivverstärker (z. B. 30 Hz entsprechend Drehzahl, 1 Hz Bandbreite, 10^4fache Verstärkung) und Ausführung liegen die Meßbereiche solcher Instrumente zwischen 10^{-3} A/m und $5 \cdot 10^6$ A/m. Die erreichbaren Meßgenauigkeiten $\Delta H/H$ liegen je nach Gerätetyp zwischen $\approx 10^{-2}$ und 10^{-4}. Es gibt auch Meßköpfe mit z. B. zwei senkrecht aufeinander stehenden Spulen zur gleichzeitigen Messung von zwei Feldkomponenten.

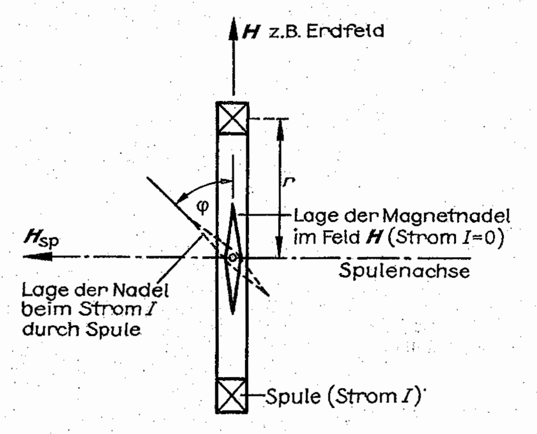


Abb. 7.4.2–3 Tangentenbussole, Magnetnadel im Mittelpunkt der Spule

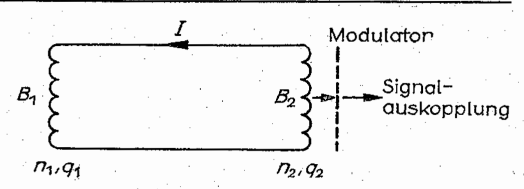


Abb. 7.4.2–4 Prinzip eines Flußtransformators

Fluxmeter dienen zur unmittelbaren Messung der magnetischen Induktion $B = \Phi/A$. Φ bezeichnet den magnetischen Fluß durch die Fläche A (gemessen in Weber Wb, 1 Wb = 1 Vs). Als Detektoren verwendet man z. B. stark gedämpfte Galvanometer, elektronische Integratoren oder supraleitende Quanteninterferometer (SQUID). Für den praktischen Gebrauch wird die Empfindlichkeit des Fluxmeters in Vs (bzw. Teile davon, meist nVs bis mVs) angegeben. Mit einfachen Zeigerinstrumenten (Drehspulfluxmeter) ergeben sich Meßbereiche zwischen $\approx 10^{-2}$ Vs bis 10^{-4} Vs.
Supraleitende Fluxmeter. Im supraleitenden Zustand (SL) eines Leiters verschwindet dessen Ohmscher Widerstand und der gesamte Fluß Φ eines Stromkreises, der z. B. aus zwei Spulen bestehen kann, bleibt konstant. Darauf beruht das Prinzip eines *Flußtransformators* (bzw. Feldtransformators). Schaltet man zwei SL-Spulen nach Abb. 7.4.2–4 zusammen und ist das Verhältnis Länge/Radius der Spulen groß gegen eins, so transformieren sich die magnetischen Induktionen nach $n_1 q_1 B_1 = n_2 q_2 B_2$. Die $n_{1,2}$ bzw. $q_{1,2}$ sind die Windungszahlen und die Querschnitte der Spulen. Damit kann man ein schwaches Feld B_2, das eine Spule mit großen n und q durchsetzt in ein starkes Feld B_1 in einer anderen Spule mit kleinen n und q transformieren und somit einfacher und genauer messen. Der SL-Strom kann durch kurzzeitige Erwärmung eines Teils des Stromkreises über den Sprungpunkt des betreffenden Supraleiters ausgeschaltet werden. Die Auskopplung des mit dem SL-Strom verbundenen magnetischen Flusses aus den Spulen ist mittels geeigneter Schaltungselemente (Modulatoren) möglich.
Zum Nachweis sehr schwacher Magnetfelder oder auch sehr kleiner Änderungen starker Felder bzw. zum Nachweis sehr kleiner Magnetisierungen $M = \chi \cdot H$ und/oder deren Änderungen verwendet man heute in zunehmendem Maße Squids. Der die Fläche eines solchen Bauelements durchsetzende magnetische Fluß muß ein *Flußquant* Φ_0 oder ein ganzzahliges Vielfaches davon sein, wobei $\Phi_0 = h/2e \approx 2 \times 10^{-15}$ Vs. Mittels geeigneter Elektronik sollten noch $\approx 0{,}02\, \Phi_0$ stabil nachweisbar sein, es lassen sich demnach (theoretisch) $5 \cdot 10^{-17}$ Vs auflösen. Mit einer angenommenen Fläche von 1 mm² würde man einen kleinsten Wert von $B = 5 \cdot 10^{-11}$ T nachweisen können. Das ist etwa der millionste Teil der Horizontalkomponente des Erdfeldes. Man kann mit modernen Geräten dieser Art z. B. auch die durch Gehirnströme erzeugten sehr schwachen Magnetfelder messen.
Magnetometer mit Gradiometeranordnung verwendet man zur Bestimmung der magnetischen Suszeptibilität $\chi = M/H$ bzw. der Magnetisierung von magnetischen Mate-

rialien in Verbindung mit SQUIDS. Eine prinzipielle Anordnung zur χ-Messung zeigt Abb. 7.4.2–5. Beim Bewegen der Probe P auf der Achse der Spulen S_1 und S_2 entsteht das skizzierte Signal, das proportional zur Flußänderung durch das Spulenpaar ist. S_1 und S_2 tragen gegenläufigen Wicklungssinn zur Unterdrückung von Fremdeinflüssen. Man kann auch die Probe an einem festen Ort lassen und die Temperatur ändern, das entstehende Signal liefert dann die Abhängigkeit der Suszeptibilität von der Temperatur. Eine besonders empfindliche Anordnung erhält man mit supraleitenden Spulen und Nachweis des Signals mittels SQUID. Das zur Messung von χ erforderliche Magnetfeld H erzeugt die Spule L, sie wird meist als Supraleitungsmagnet ausgeführt mit Induktionen bis über 10 T.

Magnetfeldmessung durch galvanomagnetische Effekte. *Hall-Sonde.* Wird nach Abb. 7.4.2–6 ein dünnes Halbleiterplättchen der Dicke d vom Strom I durchflossen und steht ein Magnetfeld B senkrecht zur Plättchenebene, so entsteht eine Spannung $U_H = R_H IB/d$ (Hall-Spannung). Der Hall-Koeffizient R_H ist materialspezifisch und beträgt für Ge z. B. $R_H = 3{,}5 \cdot 10^{-2}\,\mathrm{m^3/As}$. Mit $B = 0{,}1\,\mathrm{T}$, $I = 0{,}01\,\mathrm{A}$, $d = 0{,}1\,\mathrm{mm}$ und obigem Wert für R_H folgt $U_H = 0{,}35\,\mathrm{V}$ (Leerlaufspannung).

Widerstandssonde. Einige Metalle zeigen im Magnetfeld eine starke Abhängigkeit des elektrischen Widerstands. Als Leitermaterial wird meist Bismut (Bi) verwendet.

Thermische Effekte zur Messung von magnetischen Wechselfeldern. Infolge von *Wirbelströmen* erwärmen sich Leiter in Wechselfeldern. Dieser Effekt ist frequenzabhängig und es entstehen erhebliche Feldstörungen. Aus der Strahlungsleistung je Flächeneinheit eines elektromagnetischen Feldes W erhält man $H = \sqrt{2W/(\mu_0 c)}$. Die Strahlungsleistung je Fläche läßt sich meist thermisch messen.

Resonanzmethoden. Darunter versteht man Meßverfahren zur Bestimmung der magnetischen Feldstärke oder der Flußdichte, bei denen die Energieabsorption eine durch das

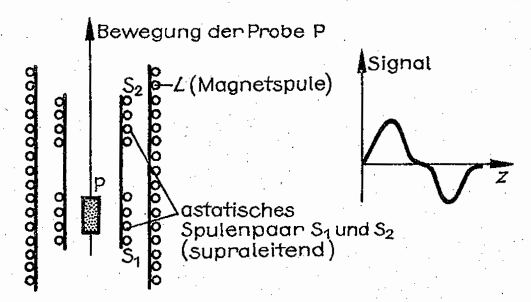


Abb. 7.4.2–5 Prinzip eines Magnetometers zur χ-Messung

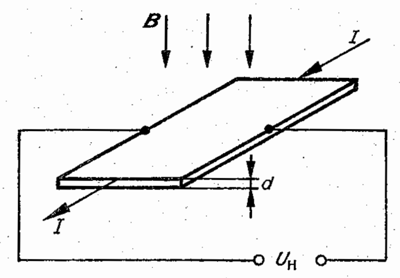


Abb. 7.4.2–6 Prinzip der Hall-Sonde

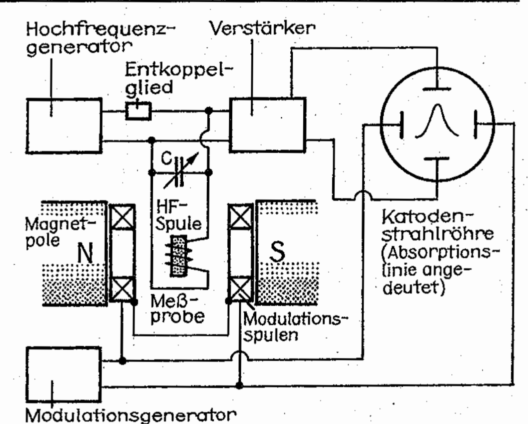


Abb. 7.4.2–7 Prinzipschaltung zur Messung der magnetischen Kernresonanz (NMR)

Magnetfeld beeinflußte Resonanz zeigt (*Zeeman-Effekt*). Als Beispiel diene die magnetische Kernresonanz (NMR).

Eine große Zahl von Atomkernen hat einen Eigendrehimpuls (Spin) und somit ein permanentes magnetisches Dipolmoment μ mit dem Betrag $\mu = \gamma\hbar\sqrt{I(I+1)}$. Hierbei bezeichnet I die Kernspinquantenzahl und γ das gyromagnetische Verhältnis des betreffenden Kernes. Die Atomkerne liefern infolge ihrer magnetischen Momente einen Beitrag $\chi_I = \mu_0 N\gamma^2\hbar^2 I(I+1)/(3kT)$ zur Gesamtsuszeptibilität χ der betreffenden Meßprobe, wobei allgemein $|\chi|_I \ll |\chi|$. (N = Anzahl der Kerne je Volumeneinheit, k = Boltzmann-Konstante, T = absolute Temperatur).

Die Wechselwirkungsenergie eines Kernmomentes μ mit einem äußeren Magnetfeld $\boldsymbol{B}_0$ beträgt $E_m = -\gamma\hbar m B_0$, wobei die magnetische Quantenzahl m die Werte $m = -I, -I+1, \ldots +I$ durchläuft. Übergänge zwischen benachbarten Energieniveaus ($\Delta m = \pm 1$) entsprechen Energiedifferenzen $\Delta E = \hbar\omega_0 = h\nu_0 = |E_{m+1} - E_m| = \gamma\hbar B_0$. Diese Energie wird einem zu $\boldsymbol{B}_0$ senkrecht stehendem Magnetfeld $\boldsymbol{B}_1$ der Frequenz $\omega_0 = \gamma B_0$ entzogen. Das ist die Resonanzbedingung der NMR. Der experimentelle Nachweis der Resonanzabsorption ist mit stationären Verfahren oder Impulsmethoden möglich. Für die Zwecke der Messung der magnetischen Feldstärke bzw. Flußdichte werden (wegen des geringeren Aufwandes) stationäre Methoden bevorzugt. Als Probesubstanz verwendet man aus Gründen der Empfindlichkeit vorzugsweise Protonen in Flüssigkeiten (meist Wasser). Einem selbstschwingenden oder fremdangesteuerten Schwingkreis wird bei der Resonanzfrequenz $\omega_0 = \gamma B_0$ Energie entzogen. Infolgedessen ergibt sich eine geringe Abnahme der Schwingkreisspannung im Bereich von $10^{-5}\,\mathrm{V}$. Zum besseren Nachweis moduliert man das Magnetfeld $\boldsymbol{B}_0$ (das bestimmt werden soll) mittels zweier kleiner Modulationsspulen um einen kleinen Wert, der etwas größer als die Linienbreite des Resonanzsignals ist. (Man kann auch die Frequenz ω modulieren.) Die zum Teil sehr schmale natürliche Linienbreite (z. B. von Protonen in Flüssigkeiten) wird für den genannten Zweck je nach erforderlicher Genauigkeit durch geringe Zugabe von paramagnetischen Ionen auf Werte zwischen $\Delta B = 10^{-6}$ bis $10^{-4}\,\mathrm{T}$ gebracht, d. h. das Feld wird maximal um $2 \cdot 10^{-4}\,\mathrm{T}$ moduliert. Durch das periodische „Überfahren“ der Resonanz-

stelle erhält man das Resonanzsignal als Wechselspannung im Takt der Modulationsfrequenz, wodurch eine bequeme Verstärkung des NMR-Signals möglich ist. Die Genauigkeit des Verfahrens wird bestimmt durch die Frequenzmessung (10^{-6} oder besser möglich) sowie Lage und Breite der Resonanzlinie. Die Werte γ_I bzw. μ_I sind zum Teil bis zu mindestens sechs Ziffern bekannt. Das NMR-Verfahren ergibt heute die genauesten absoluten Werte von Magnetfeldern. Die Abb. 7.4.2–7 zeigt eine prinzipielle Anordnung zur Messung der NMR.

Magnetfeldstabilisatoren arbeiten nach dem gleichen Prinzip. Mittels NMR kann man auch schwache Felder (etwa Erdfeld) messen. Dazu verwendet man Proben von 0,2 bis 0,5 Liter Wasser und eine entsprechend große Spule. In einem Magnetfeld von $\approx 10^{-2}$ T mit der Richtung senkrecht zu dem zu messenden Feld erreicht die makroskopische Kernmagnetisierung der Probe nach einer bestimmten Zeit (der *longitudinalen Relaxationszeit* T_1) ihren Gleichgewichtswert. Danach schaltet man das erwähnte Hilfsfeld plötzlich ab und als Folge davon präzediert die Magnetisierung um die Richtung des zu messenden Feldes (z. B. Erdfeld, B_E) mit $\nu_0 = (\gamma/2\pi)B_E \approx 2$ kHz.

Tab. 7.4.3–1 Meßgrößenumformer mit Erzeugung einer Urspannung

Meßgrößenumformer	physikalische Grundlage	natürliches Eingangssignal	typische Anwendungen
Fotoelement	Sperrschichtfotoeffekt	Bestrahlungsstärke	Beleuchtungs- und Belichtungsmesser
Thermoelement	Thermoelektrischer Effekt	Temperatur	Temperaturmessung, Strahlungsmessung
piezoelektrischer Wandler	piezoelektrischer Effekt	Kraft	Druck-, Beschleunigungs-, Schichtdickenmessung
Induktionsumformer	Induktionsgesetz	Geschwindigkeit	Tachometer, SQUID, induktive Durchflußmesser, elektromagnetisches Schallplattenabtastsystem
Galvanomagnetischer Wandler	Lorentz-Kraft	magnetische Induktion	Hall-Sonde
elektrochemische Zelle	Elektrodenpotentiale	Ionenkonzentration	elektrische *p*H-Wert-Messung
elektrostatischer Wandler	Kontaktspannung	Teilchenkonzentration	Staubgehaltsmessung in Gasen
Rauschthermometer	thermische Fluktuation des elektrischen Stromes	Temperatur	Temperaturmessung

Zur Bestimmung des Magnetfeldes an der Oberfläche von Sternen (o. ä. Objekten) kann man ebenfalls den Zeeman-Effekt nutzen. Man mißt die Aufspaltungen der betreffenden Spektrallinien sowie deren Polarisationszustand und ermittelt daraus das Magnetfeld.

7.4.3. Elektrische Messung nichtelektrischer Größen

7.4.3.1. Umformung nichtelektrischer in elektrische Meßgrößen. Bei der Messung nichtelektrischer Größen G_n (z.B. Länge, Kraft, Temperatur) ist es oft günstig, diese auf elektrische Größen G_e (z.B. Spannung, Strom, Ladung) abzubilden. Eine solche Umformung wird durch einen funktionellen Zusammenhang $G_e = f(G_n)$ beschrieben und ist ein Beispiel einer *Meßgrößenumformung*. Die Information über die zu messende nichtelektrische Größe ist dann in einem elektrischen Signal codiert und damit der elektronischen Weiterverarbeitung leicht zugänglich und für eine Fernübertragung geeignet. Meßfühler, die eine Meßgrößenumformung bewerkstelligen, werden Meßgrößenumformer, -wandler oder -geber genannt.

Auf einen Meßgrößenumformer wirken im Meßprozeß viele physikalische Größen ein. Meist interessiert aber nur die zu messende Größe und alle anderen Einwirkungen werden als

Tab. 7.4.3–2 Widerstandsgeber

physikalische Grundlage	natürliches Eingangssignal	typische Anwendungen
Abhängigkeit der elektrischen Leitfähigkeit vom Eingangssignal G_n	Temperatur	Widerstandsthermometer, Pirani-Vakuummeter, Bolometer
	Bestrahlungsstärke	Fotowiderstand, (gasgefüllte) Fotozelle
	mechanische Deformation	Dehnungsmeßstreifen
	Druck	Messung hoher und höchster Drücke
	Ionenkonzentration	elektrolytische Leitfähigkeitszelle, Feuchtemessung
	ionisierende Strahlung	Proportionszählrohr
	Magnetfeld	Magnetwiderstand
Abhängigkeit der strombegrenzenden Eigenschaft des Wandlers vom Eingangssignal $I_s = I_s(G_n)$	Bestrahlungsstärke	Fotodiode, Fototransistor, (Vakuum-)Fotozelle (mit SEV)
	(Ionen-)Konzentration	Flammenionisationsdetektor
	ionisierende Strahlung	Ionisationskammer, Geiger-Müller-Zählrohr, Halbleiterdetektor
Abhängigkeit des Widerstandes vom Stromweg	Länge	Schiebewiderstand, Wendelpotentiometer

Störungen aufgefaßt, deren Einflüsse weitgehend eliminiert werden müssen. Die zu messende Größe ist das natürliche Eingangssignal des *Wandlers*. Das natürliche Ausgangssignal (natürliches Abbildungssignal, vgl. 7.2.3.2.) der hier betrachteten Wandler liegt in elektrischer Form vor. Es kann unter Ausnutzung verschiedenartigster Naturgesetze erzeugt werden. Meßgrößenumformer lassen sich nach unterschiedlichen Gesichtspunkten ordnen. Erfolgt die Gliederung nach der zu messenden Größe, unterscheidet man thermoelektrische, optoelektrische, akustoelektrische, mechanoelektrische und andere Wandler. Die Wirkungsprinzipien dieser Wandler sind in den Abschnitten über die Messung der jeweiligen Größen beschrieben.
Im folgenden wird eine Übersicht über gebräuchliche Eingangswandler gegeben, wobei die der Meßgrößenumformung zugrunde liegenden physikalischen Wirkprinzipien im Vordergrund stehen.

7.4.3.2. Meßgrößenumformer mit Erzeugung einer Urspannung. Bei dieser Art von Wandlern entsteht aus dem Eingangssignal G_n durch Energieumformung eine elektrische Urspannung $U_0(G_n)$. Diese EMK (bzw. der daraufhin fließende Strom) ist eindeutig mit der Eingangsgröße G_n verknüpft und kann mit den in 7.4.1. beschriebenen Verfahren gemessen werden.
In Tab. 7.4.3–1 sind die wichtigsten dieser Wandler aufgeführt. Als Meßgrößen können neben den in der Spalte „natürliches Eingangssignal" genannten auch mit ihnen verknüpfte Größen dienen. Beispielsweise sind piezoelektrische Wandler auch zur Messung von Druck, Beschleunigung, Temperatur, Masse oder Längenänderung geeignet. In der letzten Spalte sind einige gebräuchliche Anwendungen genannt, die bei weitem nicht erschöpfend sind.

7.4.3.3. Meßgrößenumformer durch Beeinflussung elektrischer Größen. Hierbei entsteht das natürliche Abbildungssignal der nichtelektrischen Eingangsgröße G_n durch deren Einfluß auf einen sich in einem elektrischen Stromkreis befindlichen (i. allg. komplexen Widerstand $Z = Z(G_n)$. Je nach Charakter von Z unterscheidet man Ohmsche, induktive und kapazitive Geber. Die durch G_n verursachte Widerstandsänderung kann mit den in 7.4.1.3. beschriebenen Verfahren gemessen werden.

Resistive Meßgrößenumformer. In Tab. 7.4.3–2 sind häufig verwendete Widerstandsgeber und charakteristische Anwendungen erfaßt. Diese Wandler können grob in drei Gruppen eingeteilt werden. Während die Wandler der ersten Gruppe (Leitfähigkeitsänderung) vorwiegend in einem Ohmschen (linearen) Bereich der Strom-Spannungs-Charakteristik betrieben werden, arbeiten die Wandler der zweiten Gruppe (Strombegrenzung) in einem Bereich, wo der Strom nahezu unabhängig von der angelegten Spannung ist. Dieser Sättigungsstrom I_s wird dann im wesentlichen durch die Eingangsgröße G_n bestimmt.
Bei den Wandlern der dritten Gruppe wird durch mechanischen Eingriff (meist Drehbewegung) der Widerstand verändert.

Kapazitive Meßgrößenumformer. In Tabb. 7.4.3–3 sind typische Einsatzgebiete für kapazitive Meßgrößenumformer aufgeführt. Die relative Änderung der Kapazität eines Plattenkondensators bei Veränderung der Fläche A, des Plattenabstandes d und der Dielektrizitätskonstanten ε ergibt sich zu $\Delta C/C = \Delta A/A - \Delta d/d + \Delta\varepsilon/\varepsilon$ und führt auf die in der Tab. verwendete Einteilung.
Die größte Bedeutung haben die kapazitiven Wandler der zweiten Gruppe (Abstandsänderung). Mit modernen Herstellungstechnologien kann der Plattenabstand d bis auf 5 bis 10 µm verringert werden und die Empfindlichkeitsschwelle liegt bei 10^{-14} m. Ein weiterer Vorteil ist das im Vergleich zu resistiven und induktiven Wandlern geringe Rauschen und die fehlende Erwärmung. Kapazitive Meßgrößenumformer werden deshalb auch als höchstempfindliche Wandler in der Forschung und Technik eingesetzt.

Tab. 7.4.3–3 Kapazitive Meßgrößenumformer

Eingangssignal beeinflußt	Meßgröße
Fläche A	Winkel- und Längenänderung ($\gtrsim 1$ mm)
Abstand d	Schallgrößen (Kondensatormikrofon), Längenänderung (kapazitive Mikrometer), Kraft, Druck
Dielektrizitätskonstante	Dicke von Folien, Behälterfüllstand, Wassergehalt von Stoffen

Induktive Geber. Im engeren Sinn sind induktive Geber Induktivitäten, deren Wechselstromwiderstand durch die Eingangsgröße G_n beeinflußt wird. Induktive Geber gehören zu der großen Klasse der elektromagnetischen Umformer, die allgemein auf der Beeinflussung der magnetischen Feldenergie durch die Eingangsgröße beruhen.
Die wichtigsten Anwendungen der elektromagnetischen Umformer sind die folgenden:
- Spannungs- bzw. Stromtransformation (Meßtransformatoren und induktive Spannungsteiler),
- elektromechanische Strom-Kraft-Umformung (elektromagnetische Meßgeräte),
- Induktionsumformer (Magnetfeld- und Geschwindigkeitsmessung),
- induktive Geber zur Messung nichtelektrischer Größen, die die Geometrie des Wandlers beeinflussen,
- magnetoelastische Umformer.

Tab. 7.4.3–4 Induktive Geber

Abhängigkeit der Induktivität	Meßgröße	Anwendungsbeispiel
Lage eines (ferromagnetischen) Ankers	Längenänderung, Schichtdicke	induktive Mikrometer, Messung von Blechdicken
magnetischer Widerstand (Reluktanz) eines Magnetjoches	Kraft	magnetische Waagen
elastische Deformation des Ferromagnetikums	Kraft	magnetoelastische Kraftmeßdose
paramagnetische Suszeptibilität	Temperatur ($\lesssim 4$ K)	Kernthermometer

In Tab. 7.4.3–4 sind Beispiele für die beiden letztgenannten Einsatzbereiche der elektromagnetischen Umformer aufgeführt. Zur Erhöhung der Empfindlichkeit werden häufig *Differentialspulen* eingesetzt, bei denen durch die Eingangsgröße zwei Induktivitäten gegensinnig beeinflußt werden.

7.5. Optische Instrumente und Messung optischer Größen

Optische Instrumente dienen der Erfassung, Übertragung und Darstellung von Informationen mittels *optischer Strahlung*. Darunter wird die elektromagnetische Strahlung im Wellenlängenbereich von >10 nm (Röntgen-Gebiet) bis 1 mm (Mikrowellengebiet) verstanden (Tab. 7.5.0–1).

Tab. 7.5.0–1 Spektrale Unterteilung der optischen Strahlung

Wellenlänge	Spektralbereich	gebräuchliche Abkürzung
10...200 nm	Vakuum-Ultraviolett	VUV
200...400 nm	Ultraviolett	UV
400...760 nm	sichtbares Gebiet	VIS (von engl. visible = sichtbar)
760...2,5 μm	nahes Infrarot	nahes IR
2,5...50 μm	mittleres Infrarot	mittleres IR
50 μm...1 mm	fernes Infrarot	FIR

Abbildende optische Instrumente erzeugen von einem Objekt ein virtuelles oder reelles Bild, das vom Auge direkt oder nach fotografischer bzw. elektronischer Zwischenspeicherung und Verarbeitung betrachtet werden kann.
Optisch-physikalische Meßinstrumente werden eingesetzt zur Vermessung der Parameter optischer Strahlung (Wellenlänge, Frequenz, Polarisationszustand usw.), zur Untersuchung der Wechselwirkung elektromagnetischer Strahlung mit Materie (Absorption, Streuung, Fluoreszenz, Änderung des Polarisationszustandes u. ä.), für die Bestimmung von Stoffeigenschaften sowie zur optischen Messung nichtoptischer Größen (z. B. Länge, Winkel, Geschwindigkeit, geometrische Form und Formveränderung, elektrische und magnetische Größen, Temperatur). Moderne optisch-physikalische Meßinstrumente weisen einen hohen Integrationsgrad an optischen, feinmechanischen und elektronischen Bauelementen und Baugruppen auf.

7.5.1. Abbildende optische Instrumente

7.5.1.1. Grundlagen. Der Strahlengang in abbildenden optischen Instrumenten wird durch die Gesetze der *geometrischen Optik* beschrieben. Ein genaueres Verständnis der Wirkungsweise und der Leistungsgrenze optischer Instrumente erfordert jedoch die Berücksichtigung der *Wellennatur* der optischen Strahlung. Die Eigenschaften realer optischer Systeme werden wesentlich durch die Begrenzung der Strahlenbündel und durch die Abbildungsfehler der Linsen sowie anderer optischer Bauelemente im Strahlengang bestimmt.
In optischen Instrumenten für visuelle Beobachtung ist das menschliche Auge direkt als Sensor integriert. Es stellt einen hochintegrierten Strahlungsempfänger mit einem abbildenden optischen System variabler Brennweite und Blendenöffnung dar. Die Parameter und Toleranzen optischer Instrumente für visuelle Beobachtung müssen daher denen des Auges angepaßt werden.
Die wichtigste Grundgröße eines optischen Systems ist die *Brennweite*. Bei einfachen dünnen Linsen werden Brennweite, Objekt- und Bildabstand von der Mittelebene aus gerechnet. Bei dicken Linsen ist es zweckmäßig, zwei zur Linsenachse senkrechte Bezugsebenen, die *Hauptebenen*, einzuführen. Von der objektseitigen *Hauptebene H* werden die objektseitige Brennweite f und der Objektabstand a gezählt, entsprechend von der bildseitigen *Hauptebene H'* die bildseitige Brennweite f' und der Bildabstand a'. Für die Berechnung optischer Systeme hat es sich als günstig erwiesen, gerichtete Strecken einzuführen, wobei die Strecken in Lichtausbreitungsrichtung positiv, entgegen der Lichtrichtung negativ gezählt werden. Sammellinsen haben eine positive (bildseitige) Brennweite f'. Bei einem zentrierten Linsensystem läßt sich rein formal der einem Objektpunkt zugeordnete Bildpunkt mit Hilfe von zwei der drei ausgezeichneten Strahlen (achsenparalleler Strahl, Brennpunktstrahl, Strahl durch die Knotenpunkte) konstruieren. Befindet sich auf beiden Seiten der Linse Luft, so fallen die *Knotenpunkte* mit den Schnittpunkten der Hauptebenen mit der optischen Achse zusammen.
Vergrößerung und Abbildungsmaßstab. *Vergrößerung.* Die *scheinbare Größe* (Sehgröße) eines beobachteten Gegenstandes hängt von der Größe des Netzhautbildes im Auge ab. Es wird durch den *Sehwinkel* bestimmt, unter dem das Objekt erscheint. Der Sehwinkel wird durch die Hauptstrahlen eingeschlossen, die vom objektseitigen Knotenpunkt des Auges zu den Begrenzungen des Objektes gehen (Abb. 7.5.1–1a). Es ist üblich, die scheinbare Größe eines Gegenstandes durch den Tangens des Sehwinkels w_s anzugeben $\tan w_s = y/a_s$, wobei die *Sehweite* a_s vom Knotenpunkt zum Objekt gemessen wird. Für praktische Sehwinkel ($w_s \lesssim 10°$) gilt $\tan w_s \approx w_s$. Der Sehwinkel kann durch abbildende optische Instrumente vergrößert werden (Abb. 7.5.1–1b). Unter der *Vergrößerung* Γ (eines virtuellen Bildes) versteht man den Quotienten aus Tangens des Sehwinkels mit Instrument w_s' zum Tangens des Sehwinkels ohne Instrument $w_s \Gamma = \tan w_s'/\tan w_s$. Bei der Betrachtung sehr kleiner Gegenstände bezieht man sich auf den Sehwinkel, der bei Lage des Objektes in der deutlichen Sehweite ($a_s = 250$ mm, Normsichtweitepunkt) entsteht. Optische Instrumente zur Vergrößerung des Sehwinkels sind Lupen, Okulare, Mikroskope, Fernrohre.
Abbildungsmaßstab. Bei optischen Instrumenten, die ein reelles Bild erzeugen, wird die Vergrößerung oder Verkleinerung des Bildes gegenüber dem Objekt durch den *Abbildungsmaßstab* β als Verhältnis von reeller Bildgröße y' zur Objektgröße y angegeben $\beta = |y'|/y$ (Abb. 7.5.1–1c).
Begrenzung von Strahlenbündeln. Die durch ein optisches System übertragene Strahlungsleistung wird durch den *Öffnungs-* oder *Aperturwinkel* σ der zur Abbildung verwendeten Strahlenbündel begrenzt. Er ist durch die *Öffnungs-* oder *Aperturblende* bzw. deren Bilder festgelegt. Das objektseitige Bild der Aperturblende heißt *Eintrittspupille EP*, ihre Abbildung im Bildraum *Austrittspupille AP*. Die *Feldblende* bzw. ihre Bilder beschneiden das von einem optischen System abgebildete Gesichtsfeld. Die Bilder werden *Eintritts-* bzw. *Austrittsluke* genannt. Weiterhin beeinflussen Blenden die geometrischen Abbildungsfehler und begrenzen das durch Beugungseffekte bedingte Auflösungsvermögen des optischen Systems.
Numerische Apertur, Öffnungsverhältnis. Die *Lichtstärke* eines optischen Systems wird durch die übertragbare Strahlungsleistung ausgedrückt, die durch den objektseitigen Öffnungswinkel σ des Strahlenbündels festgelegt ist (Abb. 7.5.1–1c). $\sin \sigma$ heißt die *Apertur* des optischen Systems. Befindet sich ein Medium mit der Brechzahl $n \neq 1$ zwischen Objekt und Objektiv, so ist die *numerische Apertur*

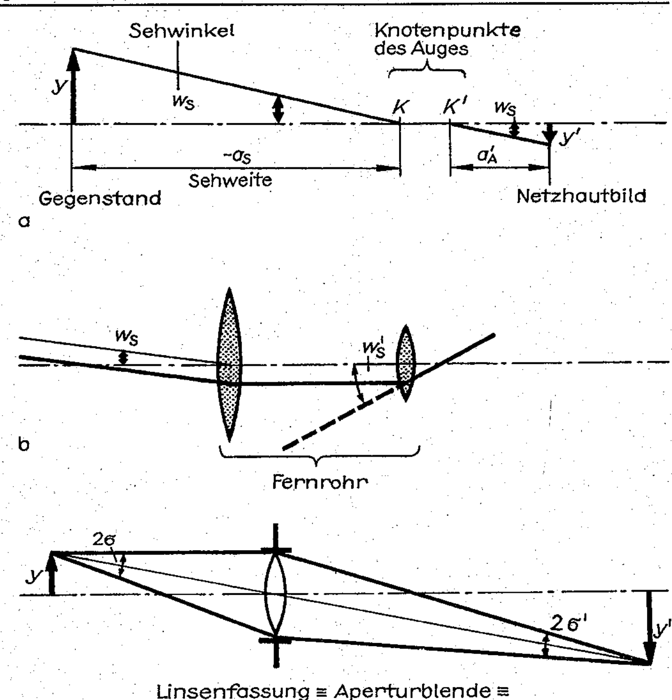


Abb. 7.5.1-1 (a) Sehwinkel ohne Instrument, (b) Sehwinkel mit Instrument, (c) Abbildung mit Strahlenbündel (Öffnungswinkel σ)

$A = n \sin \sigma$ ein Maß für die in das Objektiv eintretende Strahlungsleistung. Diese Angabe ist für die Abbildung mit weit geöffneten Bündeln (Mikroskop) sinnvoll. Werden dagegen sehr weit entfernte Objektive mit engen Bündeln abgebildet (Fernrohr), so ist es günstiger, die übertragene Strahlungsleistung durch das *Öffnungsverhältnis* $K = 1/B$ zu charakterisieren. Für ein unendlich fernes Objekt ist es definiert als Verhältnis von Durchmesser D der Eintrittspupille zur Brennweite f, $K = D/f$. Der Reziprokwert $B = 1/K$ wird *Blendenzahl* genannt. Da die Bestrahlungsstärke E in der Bildebene mit der Fläche der Eintrittspupille ($\sim D^2$) zu- und mit größer werdendem Bild (Bildfläche $\sim f^2$) abnimmt, gilt $E \sim D^2/f^2 = 1/B^2$. Öffnungsverhältnis bzw. Blendenzahl B sind für unendlich ferne Objekte mit der numerischen Apertur A verknüpft: $B = 1/(2A)$.

Abbildungsfehler (Aberrationen). Bei realen optischen Systemen werden Strahlenbündel genutzt, die außerhalb des paraxialen Gebietes liegen. Sie führen zu Abweichungen von der idealen geometrisch-optischen Abbildung, die als Bildfehler oder Aberrationen bezeichnet werden.

Chromatische Bildfehler werden durch die Wellenlängenabhängigkeit der Brechzahl verursacht. Sie werden durch Linsenkombinationen aus Gläsern unterschiedlicher Brechzahl und Dispersion (vgl. 7.5.4.1.) beseitigt. Linsensysteme, deren Farbfehler für zwei Wellenlängen korrigiert wurde, heißen *Achromate*. Sie werden *Apochromate* genannt, wenn die Korrektur für drei Wellenlängen erfolgte. Hohlspiegel haben keinen Farbfehler. Bei monochromatischer Strahlung treten die folgenden Bildfehler auf.

Öffnungsfehler (*sphärische Aberration*). Strahlen, die von einem Objektpunkt auf der Achse unter verschiedenen Öffnungswinkeln σ auf das optische System treffen, werden nicht in einem Bildpunkt vereinigt, sondern schneiden die optische Achse vor oder hinter dem idealen (Gaußschen) Bildpunkt. Die Distanz zwischen Durchstoßpunkt und idealem Bildpunkt heißt *Längsaberration*. Eine Korrektur erfolgt durch Kombination von Sammel- und Zerstreuungslinsen. Bei der Abbildung außeraxialer Objektpunkte muß zusätzlich die Abbesche *Sinusbedingung* eingehalten werden: $\beta n \sin \sigma = n' \sin \sigma'$ (σ' = Öffnungswinkel des bildseitigen Strahlenbündels, n' = Brechzahl im Bildraum, β = Abbildungsmaßstab). Optische Systeme mit erfüllter Sinusbedingung und korrigiertem Öffnungsfehler heißen *Aplanate*.

Koma (*Asymmetriefehler*). Schief einfallende Strahlenbündel, die einseitig von einer Blende beschnitten werden, erzeugen Zerstreuungsfiguren mit schweifförmigem Aussehen.

Astigmatismus. Schief zur optischen Achse einfallende Strahlenbündel werden in der Ebene, die von der Symmetrieachse des Bündels (Hauptstrahl) und der optischen Achse des Systems aufgespannt wird (Meridionalebene), und in der dazu senkrecht liegenden Ebene (Sagittal-Ebene) unterschiedlich stark gebrochen, da die Krümmungsradien der Kreisbögen, die beim Schnitt der sphärischen Linsenfläche mit den beiden Ebenen zustande kommen, unterschiedlich sind. Als Folge entstehen von einem außerhalb der Achse gelegenen Objektpunkt in zwei Bildweiten zwei aufeinander senkrecht stehende Bildlinien (astigmatische Abbildung).

Bildfeldwölbung. Der kleinste Zerstreuungskreis des bildseitigen Strahlenbündels liegt i. allg. auf einer gekrümmten Rotationsfläche. Objektive, deren Bildfläche eingeebnet wurde, nennt man *Planfeldobjektive*. Zur Korrektur von Bildfeldwölbung und Astigmatismus muß die sogenannte *Petzval-Summe* möglichst klein gemacht werden.

Verzeichnung. Wenn der Abbildungsmaßstab sich mit dem Bildwinkel ändert, kommt es zu einer „kissen"- oder „tonnenförmigen" Verzeichnung.

Eigenschaften des Auges. *Aufbau des Auges*. Die lichtempfindlichen Zellen sind in der inneren Augenhaut (*Netzhaut* oder *Retina*) angeordnet. Die Abbildung des Objektes auf die Netzhaut geschieht durch ein zusammengesetztes optisches System. Die Lichtbrechung erfolgt wegen des Überganges Luft – hochbrechendes Medium vornehmlich an der vorderen Eintrittsfläche, der Hornhaut (Brechkraft ≈ 42 Dioptrien). Die hinter der kontrahierbaren Blende (Pupille oder Iris) befindliche Augenlinse übt nur einen korrigierenden Einfluß auf die Gesamtbrechkraft aus. Die Krümmung der Linse und damit ihre Brechkraft lassen sich durch einen ringförmigen Muskel (Ziliarmuskel) vergrößern, um nahe liegende Objekte auf der Netzhaut scharf abzubilden. Dieser Vorgang wird *Akkommodation* genannt. Ein normalsichtiges Auge erzeugt im nichtakkommodierten Zustand ein scharfes Bild eines unendlich fernen Gegenstandes (*Fernpunkt*). Durch Akkommodation gelingt es kurzzeitig einem normalsichtigen jugendlichen Auge, ein Objekt noch in einer Entfernung von ≈ 100 mm scharf zu sehen (*Nahpunkt*). Ohne Ermüdung des Auges muß jedoch eine Gegenstandsentfernung von 200 bis 300 mm eingehalten werden. Die Entfernung von 250 mm wird als *deutliche Sehweite* definiert. Die maximale Änderung der Brechkraft der Augenlinse beträgt im jugendlichen Alter ≈ 10 Dioptrien. Die Akkommodation vom Nah- in den Fernbereich erfolgt in einer Zeit von ≈ 0,4 s. Akkommodation und Akkommodationsgeschwindigkeit nehmen mit zunehmendem Lebensalter ab. In der Regel ist im Alter über ≈ 45 Jahre ein Gegenstand in der deutlichen Sehweite für längere Zeit nicht mehr scharf zu erkennen.

Lichtempfindliches System. Die Netzhaut enthält $\approx 1{,}2 \cdot 10^8$ lichtempfindliche Zellen (Rezeptoren), in denen bei Auftreffen optischer Strahlung ein komplizierter physikalisch-chemischer Prozeß abläuft. Die Rezeptoren sind einzeln oder in Gruppen mit Nervenleitungen verbunden, die den Lichtreiz zum Gehirn leiten. Ihre Verteilung auf der Netzhaut ist unterschiedlich. Der größte Teil der Rezeptoren besteht aus Zäpfchen, die das Farbsehen vermitteln und nur bei hohen Leuchtdichten ($\gtrsim 3\,\mathrm{cd/m^2}$ entsprechend einer Strahlungsflußdichte $M_e \gtrsim 5 \cdot 10^{-3}\,\mathrm{W/m^2}$) reagieren. Das Maximum liegt für hohe Strahlungsdichten bei $\lambda = 555$ nm. Die Stäbchen sprechen auf geringe Leuchtdichten ($\ll 3\,\mathrm{cd/m^2}$) an (*Dämmerungssehen*). Sie können aber keine großen Farbunterschiede wahrnehmen. Die Stäbchen sind in den äußeren Bereichen der Netzhaut vorherrschend. Die minimal wahrnehmbare Bestrahlungsstärke beträgt im spektralen Maximum $\approx 6 \cdot 10^{-13}\,\mathrm{W/m^2}$. Das entspricht bei einer Pupillenfläche von $5 \cdot 10^{-5}\,\mathrm{m^2}$ einer Strahlungsleistung von $\approx 3 \cdot 10^{-17}$ W entsprechend einem Fluß von 100 Photonen $\mathrm{s^{-1}}$ oder einem Lichtstrom von $\approx 2 \cdot 10^{-14}$ Lumen. Unter besonders günstigen Umständen vermag das Auge noch Lichtblitze von wenigen Photonen wahrzunehmen.

Sehschärfe (*Auflösungsvermögen*). Die Zäpfchen sind vornehmlich im zentralen Teil der Netzhaut (Fovea centralis) konzentriert, wo jedes Zäpfchen mit einer Nervenfaser verbunden ist. Ihre Durchmesser betragen ≈ 1,5 bis 5 µm. Zwei punktförmige Objekte werden vom Auge getrennt wahrgenommen, wenn ihre Bilder auf zwei verschiedene Zäpfchen fallen. Für Zäpfchen in der Fovea entspricht das bei einer Brennweite des normalen Auges von 15 mm einer Winkelauflösung von ≈ 0,3 mrad ≙ 1′. Der Winkelbereich großen Auflösungsvermögens, d. h. großer *Sehschärfe*, liegt symmetrisch zur Fovea und beträgt nur wenige Bogengrade. Bereits innerhalb eines Kegels von ± 2° fällt das Auflösungsvermögen auf die Hälfte und innerhalb ± 10° auf ein Fünftel seines Maximums ab. Zwei parallele, sich ungefähr fortsetzende Striche können voneinander getrennt werden, wenn ihr Winkelabstand ≈ 0,05 mrad ≙ 10″ beträgt (*Noniussehschärfe*). Die Winkelauflösung des Auges begrenzt die sinnvolle (förderliche) Vergrößerung optischer Instrumente für visuelle Beobachtung. Der geringe Winkelbereich für die größte Sehschärfe hat zur Folge, daß die Beobachtung mit Mikroskopen und Fernrohren in der Regel mit bewegtem Auge bzw. durch Drehen des Kopfes erfolgt.

Adaption. Das Auge ist in der Lage, Bilder in einem sehr großen Leuchtdichtebereich von $\approx 2 \cdot 10^{-6}$ bis $\approx 2 \cdot 10^5\,\mathrm{cd\,m^{-2}}$ zu verarbeiten. Diese Anpassung des Auges an unterschiedliche Objekthelligkeiten bezeichnet man als *Adaption*. Sie beruht auf einem komplexen physikalisch-chemischen und physiologischen Vorgang. Die Zeit für die Anpassung des Zäpfchensehens beträgt ≈ 3 bis 5 min (*Sofortadaption*), die des Stäbchensehens ≈ 8 bis 10 min (*Daueradaption*). Für die *Dunkeladaption* werden ≈ 40 min benötigt. Die Schwächung des Lichtstromes durch die als Blende wirkende, kontrahierbare Pupille (Durchmesser 1,5 bis 8 mm) ist für die Adaption nur von geringer Wirkung. Bei Pupillendurchmesser > 5 mm wird jedoch durch chromatische und sphärische Aberration das Auflösungsvermögen vermindert. Der Durchmesser des aus einem optischen Instrument austretenden Lichtbündels sollte daher 4 bis 5 mm nicht übersteigen.

Flimmer- und Verschmelzungsfrequenz. Periodische Lichtreize kurzer Dauer werden vom Auge bei geringer Folgefrequenz getrennt wahrgenommen. Bei höherer Frequenz ($\approx 3\,\mathrm{s^{-1}}$ bis $5\,\mathrm{s^{-1}}$ bei mittlerer Leuchtdichte) tritt Flimmern auf. Oberhalb der *Verschmelzungsfrequenz* ($\approx 30\,\mathrm{s^{-1}}$) entsteht ein kontinuierlicher Helligkeitseindruck, da das Auge zeitlich mittelt (*Talbotsches Gesetz*).

Wahrnehmung von Leuchtdichteunterschieden. Zahlreiche optische Meßverfahren beruhen auf dem Leuchtdichteabgleich zweier benachbarter Beobachtungsfelder mit den Leuchtdichten L und $L + \Delta L$. Die Empfindlichkeit des Auges, einen Leuchtdichteunterschied ΔL wahrzunehmen, hängt ab von der *Unterschiedsempfindlichkeit* $L/\Delta L$. Sie beträgt im Leuchtdichteintervall von $10\,\mathrm{cd\,m^{-2}} \lesssim L \lesssim 10^3\,\mathrm{cd\,m^{-2}}$ unter optimalen Bedingungen und adaptiertem Auge $L/\Delta L \approx 40$ bis 100 (Abb. 7.5.1–2).

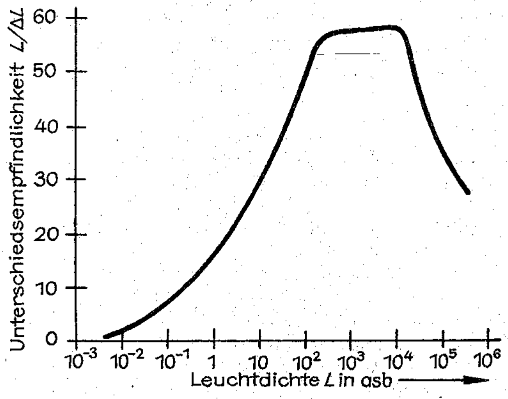


Abb. 7.5.1–2 Wahrnehmung von Leuchtdichteunterschieden (1 Apostilb = 1 asb = $1/\pi\,\mathrm{cd\,m^{-2}}$)

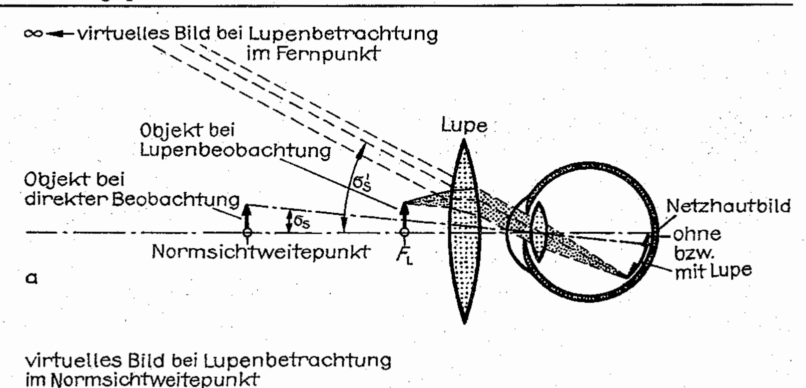


Abb. 7.5.1–3 Vergrößerung der Lupe

Wahrnehmungsgeschwindigkeit. Das Auge ist in der Lage, Bewegungen eines Objektes im Winkelgeschwindigkeitsbereich von (15 bis 30)'s^{-1} bis zu (150 bis 350)°s^{-1} wahrzunehmen. Bei Vergleich mit einem ruhenden Gegenstand können noch Objektbewegungen von (1 bis 2)'s^{-1} erkannt werden. Die Werte hängen stark von der Leuchtdichte ab.

7.5.1.2. Lupen. Die *Lupe* ist das einfachste optische Instrument zur Vergrößerung naher Objekte. Bei geringer Vergrößerung ($\Gamma < 6$) besteht sie aus *einer* Sammellinse kleiner Brennweite $f' = -f$. Ein Gegenstand, der sich in der objektseitigen Brennebene befindet, erzeugt bei Beobachtung mit einem normalsichtigen, völlig entspannten Auge hinter der Lupe ein virtuelles Bild im Unendlichen (Abb. 7.5.1–3). Die Vergrößerung Γ_N der Lupe beträgt in diesem Fall $\Gamma_N = 250\,\text{mm}/f'$ (*Normalvergrößerung*). Wird die Lupe so verwendet, daß das virtuelle Bild in der deutlichen Sehweite erscheint, so erhöht sich die Vergrößerung Γ auf $\Gamma = \Gamma_N + 1$. Lupen höherer Vergrößerung bestehen aus mehreren Linsen, deren Abbildungsfehler weitgehend korrigiert sind. *Aplanatische Lupen* (korrigierter Öffnungsfehler, erfüllte Sinusbedingung) erlauben 6- bis 15fache Vergrößerungen. Die obere mit Lupen erreichbare Grenze liegt bei $\Gamma \approx 20$ bis 30. Mit steigender Vergrößerung nimmt das übersehbare Feld ab.

Ableselupen werden zur Erhöhung der Ablesegenauigkeit von Skalen eingesetzt (ohne Lupe kann das Auge in der normalen Ableseentfernung von 250 mm Skalenteilungen mit einer Genauigkeit von ± 0,15 mm bequem ablesen). *Meßlupen* ($\Gamma \approx 8$ bis 10) sind mit einer Halterung versehen, die eine Skale von 10 mm Länge (geteilt in 0,1 mm) aufweist, und dienen zur Messung kurzer Strecken.

Okulare sind speziell korrigierte zwei- oder mehrlinsige optische Systeme zur visuellen Betrachtung des von einem Objektiv (Fernrohr-, Mikroskop-Objektiv) erzeugten Zwischenbildes. Mit dem Okular wird das Zwischenbild vergrößert und als virtuelles Bild im „Unendlichen" abgebildet.

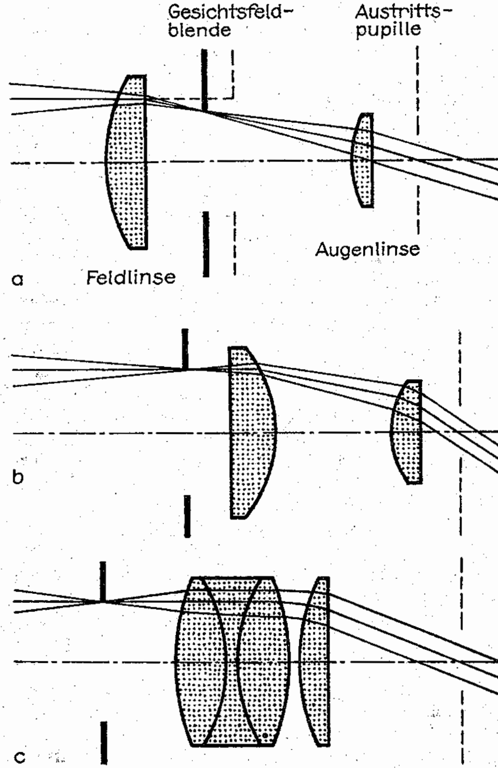


Abb. 7.5.1–4 Okulartypen, (a) Huygenssches, (b) Ramsdensches und (c) orthoskopisches Okular

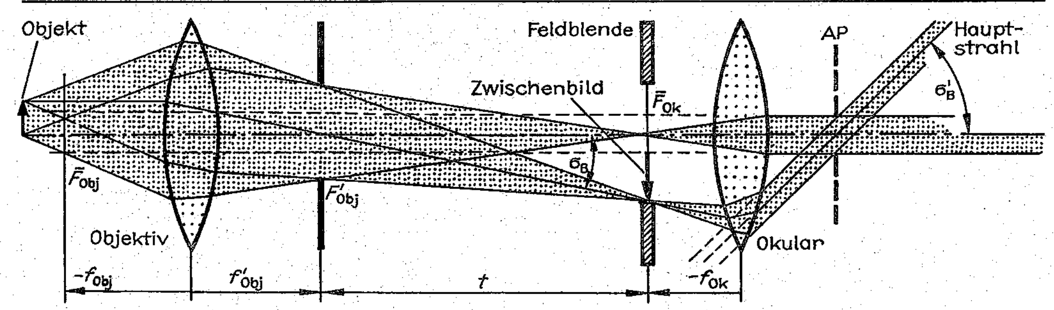


Abb. 7.5.1–5 Strahlengang im Mikroskop

Es kann dann bequem mit nichtakkommodiertem Auge beobachtet werden (Lupenfunktion des Okulars). Außerdem bildet es die Eintrittspupillen des Objektives in die Augenpupille ab, damit alle bilderzeugenden Strahlen in das Auge gelangen können.

Spezielle Okulare. (Abb. 7.5.1–4) Grundformen der Okulare ohne Bildumkehr sind das Okular mit Mittenblende (Huygenssches Okular) und das Okular mit Vorderblende (Ramsdensches Okular). Das *Huygenssche Okular* besteht aus der plankonvexen Feldlinse und einer plankonvexen Augenlinse. Die Planseiten sind dem Auge zugewandt. Das Zwischenbild entsteht zwischen den Linsen. Für Meßzwecke wird vorzugsweise das *Ramsdensche Okular* eingesetzt, dessen Zwischenbildebene vor dem Okular liegt. Sie ist leicht zugänglich, so daß dort Meßmarken oder Mikrometerteilungen angebracht werden können, die mit dem Bild gleichzeitig scharf gesehen werden. Beim oft eingesetzten *orthoskopischen Okular* verzichtet man auf die Feldlinse. Es ähnelt einem einlinsigen Okular, hat aber einen hohen Korrekturzustand. Ein häufiger Anwendungsfall ist der Einsatz als Mikrometerokular. Der wirksame Durchmesser der Feldblende des Okulars wird als *Feldzahl* bezeichnet.

Das *Gaußsche Okular* wird speziell bei Autokollimationsfernrohren eingesetzt. Es enthält eine beleuchtete Marke, die durch Autokollimation wieder in sich selbst abgebildet wird. Sie wird über eine teildurchlässige dünne planparallele Platte von der Seite beleuchtet.

7.5.1.3. Mikroskope dienen zur Vergrößerung sehr kleiner Objekte. Sie haben eine zweistufige Abbildung. Mit dem ersten sammelnden optischen System (Objektiv) wird ein reelles Zwischenbild erzeugt (Abb. 7.5.1–5), das mit einem zweiten sammelnden System (Okular oder Projektiv) nachvergrößert wird. Okulare dienen der visuellen Beobachtung. Es wird ein virtuelles Bild im „Unendlichen“ (Normalvergrößerung) oder in der deutlichen Sehweite erzeugt. Projektive erzeugen reelle Endbilder. Sie werden für die Mikrofotografie und Mikroprojektion eingesetzt. Die Güte eines Mikroskopes wird wesentlich durch die Mikroskopobjektive bestimmt. In der Regel gehören mehrere Mikroskopobjektive zu einer Objektivreihe, deren Abbildungsmaßstäbe und numerischen Aperturen abgestuft sind.

Vergrößerung. Die resultierende Brennweite f'_M eines Mikroskopes, das aus den zwei Linsensystemen mit den Brennweiten f'_{Obj} und f'_{Ok} besteht, beträgt $f'_M = f'_{Obj} \cdot f'_{Ok}/t$. Der Abstand zwischen den beiden Brennpunkten von Objektiv und Okular wird *optische Tubuslänge t* genannt. Durch geeignete Wahl der Tubuslänge kann eine Mikroskopgesamtbrennweite f'_M erzielt werden, die wesentlich kleiner als die der Brennweiten von Objektiv und Okular ist. Bei Beobachtung mit nichtakkommodiertem Auge ergibt sich als Normalvergrößerung des Mikroskopes $\Gamma_M = 250\,\text{mm}/f'_M = t \cdot 250\,\text{mm}/f'_{Obj} f'_{Ok} = \beta_{Obj}\,\Gamma_{Ok}$.

Die Gesamtvergrößerung setzt sich multiplikativ aus dem Abbildungsmaßstab des Objektivs $\beta_{Obj} = t/f'_{Obj}$ und der Vergrößerung des Okulars $\Gamma_{Ok} = 250\,\text{mm}/f'_{Ok}$ zusammen. Für typische Werte $f'_{Obj} = 4\,\text{mm}$, $f'_{Ok} = 20\,\text{mm}$ und $t = 160\,\text{mm}$ berechnet sich eine resultierende Mikroskopbrennweite $f'_M = 0{,}5\,\text{mm}$. Der Abbildungsmaßstab des Objektives ist damit $\beta_{Obj} = 40$, die Okularvergrößerung $\Gamma_{Ok} = 12{,}5$ und die Gesamtvergrößerung $\Gamma_M = 500$.

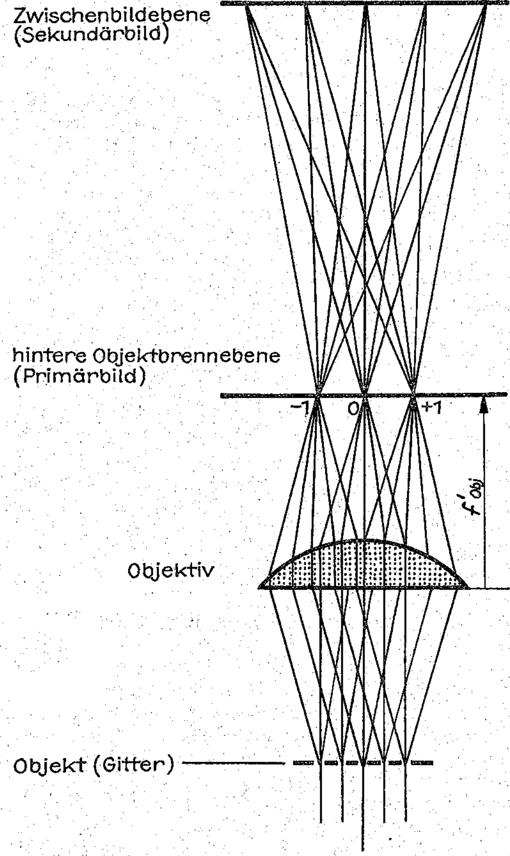


Abb. 7.5.1–6 Bildentstehung im Mikroskop nach Abbe

Bildentstehung im Mikroskop (ABBE 1873). Zum genaueren Verständnis der Bildentstehung im Mikroskop müssen die Welleneigenschaften des Lichtes und die dadurch bedingten Beugungs- und Interferenzeffekte berücksichtigt werden. Bei der Beobachtung von *nichtselbstleuchtenden Objekten* ist der Einfluß der Beleuchtungsquelle zu beachten. Zur Verfeinerung wird eine punktförmige Lichtquelle betrachtet, die genügend weit vom Objekt entfernt ist, so daß ihre Strahlung die Objektebene praktisch als ebene Welle (Parallellichtbündel) durchsetzt (Abb. 7.5.1–6). Das Objektiv erzeugt dann ein reelles Bild der Lichtquelle in der hinteren Brennebene (genauer: das Airysche Beugungsscheibchen der Lichtquelle, das durch Beugung an der Öffnung des Objektives zustande kommt). Von dort läuft das Licht als Kugelwellen auseinander und in der Zwischenbildebene erscheint das gleichmäßig erleuchtete Bild der Objektebene. Wird ein Objekt mit feinen sichtbaren Strukturen in die Objektebene gebracht, so treten bei der Durchleuchtung mit parallelem kohärenten Licht Fraunhofersche Beugungserscheinungen auf. In der bildseitigen Brennebene des Objektives entstehen dann die Beugungsfiguren des Objektes in den unterschiedlichen Ordnungen als Bilder der Lichtquelle (*Primäres Bild*). Durch Interferenz der von der Beugungsfigur ausgehenden Kugelwellen wird in der Zwischenbildebene das reelle Bild des Objektes (*Sekundäres Bild*) erzeugt. Bei Verwendung eines Strichgitters (Amplitudengitter) als Objekt können seine Beugungsbilder 0., ± 1., ± 2. ... ± n-ter Ordnung der Lichtquelle gut voneinander getrennt werden. Das Zwischenbild stimmt nur dann mit dem Objekt völlig überein, wenn alles vom Objekt abgebeugte Licht in das Objektiv gelangt und für die Bildentstehung zur Verfügung steht. Je feiner die Strukturen sind, desto größer ist der Beugungswinkel für die Nebenmaxima. Durch den konstruktiv festgelegten Öffnungswinkel des Objektives gelangt bei sehr feinen Strukturen aber nur ein Teil der Beugungsordnungen in das Objektiv. Sind die Beugungswinkel so groß, daß nur das Licht des 0. Hauptmaximums in das Objektiv eintritt, so können keine Strukturen im Bild erkannt werden.

Auflösungsvermögen. Ein Objekt, das aus zwei dicht benachbarten Gitterspalten im Abstand d_{min} besteht, wird dann noch aufgelöst, wenn mindestens die ersten Beugungsordnungen in das Objektiv gelangen (Abb. 7.5.1–7). Die *numerische Apertur* $A_{Obj} = n \sin \sigma$ muß dazu größer oder mindestens gleich dem Sinus des Beugungswinkels α des ersten Nebenmaximums sein: $A_{Obj} = n \sin \sigma \geqq \sin \alpha = \lambda / d_{min}$.
Darin bedeuten λ die Lichtwellenlänge und n die Brechzahl des Mediums zwischen Objekt und Objektiv. Der kleinste auflösbare Abstand beträgt daraus $d_{min} \geqq \lambda / A_{Obj}$. Das entstandene Bild ist jedoch nicht objekttreu.

Methoden zur Erhöhung des Auflösungsvermögens.

- Vergrößerung der numerischen Apertur A_{Obj}: Trockensysteme (Luft zwischen Objekt und Objektiv [$n = 1$]) haben eine maximale numerische Apertur $A_{Obj} = 0{,}95$ entsprechend einem Aperturwinkel $\sigma \approx 72°$. A_{Obj} kann durch Übergang zu Immersionsobjektiven vergrößert werden, bei denen sich eine Immersionsflüssigkeit zwischen Objekt und Objektiv befindet. Übliche Immersionsflüssigkeiten sind Wasser ($n = 1{,}33$), Glyzerin-Wasser ($n \approx 1{,}45$), Zedernholzöl ($n = 1{,}52$) oder Monobromnaphthalin ($n = 1{,}66$). Mit letzterer Flüssigkeit läßt sich die numerische Apertur bis $A_{Obj} = 1{,}60$ steigern. Damit können noch Strukturen mit einem Abstand $d_{min} \approx 200$ nm ($\lambda = 450$ nm) aufgelöst werden.
- Übergang zu kürzeren Wellenlängen in das UV-Gebiet.
- Durch *schräge Hellfeldbeleuchtung* kann nur eine Beugungsordnung mit dem doppelten Beugungswinkel 2α noch vom Objektiv aufgenommen werden (Abb. 7.5.1–8). Der minimal auflösbare Abstand d_{min} verkürzt sich dadurch auf $d_{min} = \lambda / 2A_{Obj}$.

Förderliche Vergrößerung. Die Strukturen des vom Okular vergrößerten Zwischenbildes werden vom Auge nur dann bequem aufgelöst, wenn sie unter einem Winkel von 2′ bis 4′ erscheinen. Daraus ergibt sich als notwendige Gesamtvergrößerung Γ_M des Mikroskopes (sogenannte *förderliche*

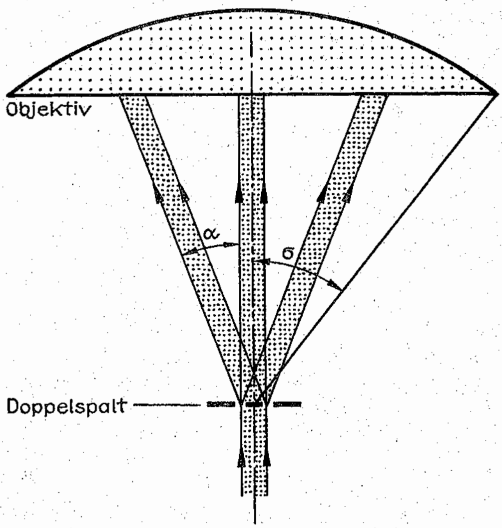


Abb. 7.5.1–7 Begrenzung des Auflösungsvermögens durch die numerische Apertur

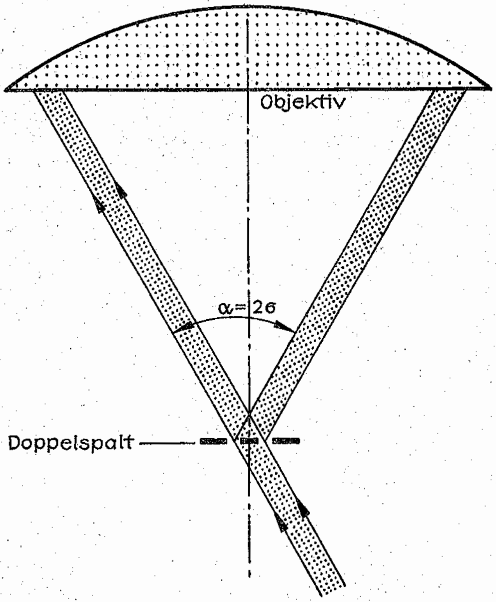


Abb. 7.5.1–8 Schräge Hellfeldbeleuchtung

Vergrößerung) $\Gamma_M = (500 \text{ bis } 1000) \cdot A_{Obj}$. Eine stärkere Vergrößerung würde keine feineren Strukturen erkennen lassen. Die stärkste sinnvolle Vergrößerung eines Lichtmikroskopes beträgt danach $\Gamma_M \approx 1600$ (für $A_{Obj} = 1{,}6$).

Schärfentiefe. Für mikroskopische Arbeiten ist es erforderlich, den Abstand Objekt – Objektiv so einzustellen, daß das Bild möglichst „scharf" erscheint. Weicht die Einstellebene von der Ebene der optimalen Schärfe ab, so nimmt der Durchmesser eines Bildpunktes zu. Die Ursachen sind sowohl geometrisch-optisch (Übergang eines Bildpunktes in einen Zerstreuungskreis, (vgl. 7.5.1.5.) als auch wellenoptisch (räumliche Ausdehnung der Beugungsfigur) bedingt. In dem Bereich der Schärfentiefe ist die Zunahme eines „Bildpunktes" so klein, daß sie vom Auge nicht wahrgenommen wird. Ohne Berücksichtigung der Akkommodation des Auges (z.B. bei mikrofotografischer Abbildung) beträgt die Tiefenschärfe

$$t_m = n((\lambda/\text{mm})/2\,A_{Obj}^2 + 0{,}15/A_{Obj}\,\Gamma_M)\ [\text{mm}].$$

Abgleich. Die optischen und mechanischen Parameter von Objektiven und Okularen werden in der Regel so dimensioniert („abgeglichen"), daß bei einem Wechsel die Tubuslänge nicht verstellt werden muß. Der Abstand Objektebene–Anlagefläche Objektiv (= Abgleichlänge) beträgt bei einer Reihe von Mikroskopherstellern bis auf Ausnahmen 45 mm.

Beleuchtungsarten. Mikroskopische Präparate sind meist nicht selbstleuchtend. Sie bedürfen daher einer zusätzlichen Beleuchtung. Durchsichtige Objekte werden mit *Durchlichtbeleuchtung*, die Oberflächen von undurchsichtigen Objekten mit *Auflichtbeleuchtung* beobachtet. Weiterhin wird zwischen Hellfeld- und Dunkelfeldbeleuchtung unterschieden. Bei der *Hellfeld-Beleuchtung* gelangen sowohl das direkte Licht als auch die Beugungsordnungen in das Objektiv. Die Objektstrukturen erscheinen dunkel auf hellem Feld. Bei der *Dunkelfeld-Beleuchtung* wird die 0. Beugungsordnung ausgeblendet. Es trägt nur das abgebeugte Licht zur Bildentstehung bei, so daß Objektstrukturen hell auf dunklem Hintergrund sichtbar werden.

Köhlersches Beleuchtungsverfahren. Es wird meist für die Durchlicht-Hellfeldmikroskopie angewandt (Abb. 7.5.1–9). Die Lichtquelle wird mittels der Kollektorlinse vergrößert in die objektseitige Brennebene der Kondensorlinse abgebildet, in der eine als Beleuchtungsaperturblende (Öffnungsblende) wirkende Irisblende angeordnet ist. Sie bestimmt die Apertur der aus dem Kondensor zur Beleuchtung des Objektes austretenden Parallellichtbündel. Der Kondensor bildet weiter die hinter dem Kollektor befindliche Leuchtfeldblende in die Objektebene ab, mit deren Hilfe die Größe des beleuchteten Objektausschnittes variiert werden kann. Die Vorteile dieses Beleuchtungssystems bestehen darin, daß das Objekt gleichmäßig mit parallelem Licht ausgeleuchtet wird, die Größe des beleuchteten Objektbereiches verändert werden kann, um das Objekt vor unnötiger Erwärmung und Überstrahlung zu schützen, und daß sich unabhängig davon die Apertur der beleuchtenden Bündel einstellen läßt.

Dunkelfeldbeleuchtung. Das einfachste Verfahren zur Ausblendung des direkten Lichtbündels ist die schräge Beleuchtung, die jedoch durch Bevorzugung einer Richtung eine Reihe von Nachteilen aufweist. Günstiger ist eine allseitige Beleuchtung mittels spezieller Dunkelfeldkondensoren, die in der Regel als Spiegelkondensoren ausgebildet sind (z.B. *Kardioidkondensor*). Auch der Einsatz einer Zentralblende in einem normalen Hellfeldkondensor ist üblich.

Auflichtbeleuchtung. Undurchsichtige Strukturen werden von der Objektseite aus beleuchtet, wobei mit dem reflektierten, gestreuten oder gebeugten Licht beobachtet wird. Wenn der Arbeitsabstand genügend groß ist, kann sich die Beleuchtungsquelle neben dem Tubus befinden. Bei Objektiven mit kleinem Arbeitsabstand erfolgt die Beleuchtung durch das Objektiv hindurch (Abb. 7.5.1–10).

Immersionssysteme. Die numerische Apertur vom Objektiv bzw. von der Beleuchtungsoptik wird durch Einbringen einer Flüssigkeit mit hoher Brechzahl zwischen Deckglas und Objektiv bzw. zwischen Kondensor und Objektträger erhöht. Wenn die Brechzahlen von Deckglas und Immersionsflüssigkeit gleich sind, spricht man von *homogener Immersion*.

Dunkelfeld und Phasenkontrast. Biologische Mikroskoppräparate (Dünnschnitte, Ausstriche) zeigen häufig nur Phasenstrukturen, die sich erst durch spezielle Färbetechniken kontrastreich von der Umgebung abheben. Ungefärbte Phasenobjekte können sichtbar gemacht werden, wenn durch Eingriffe in das primäre Bild solche Amplituden- und Phasenverhältnisse hergestellt werden, wie sie bei einem Amplitudenpräparat auftreten.

Bei der *Dunkelfeld-Beleuchtung* kann nur Licht, das durch Beugung, Reflexion oder Streuung an den Objektstrukturen in das Objektiv gelangt, zur Bildentstehung beitragen. Dadurch erscheinen auch Phasenstrukturen hell auf dunklem Grund. Für den Nachweis schwacher Phasenobjekte mit Phasenverschiebungen von nur $\Delta\varphi \approx 3° \mathrel{\hat{=}} \lambda/_{20}$ ist das *Phasenkontrastverfahren* geeignet. Bei ihm wird in der bildseitigen Brennebene des Objektives nicht nur die Amplitude

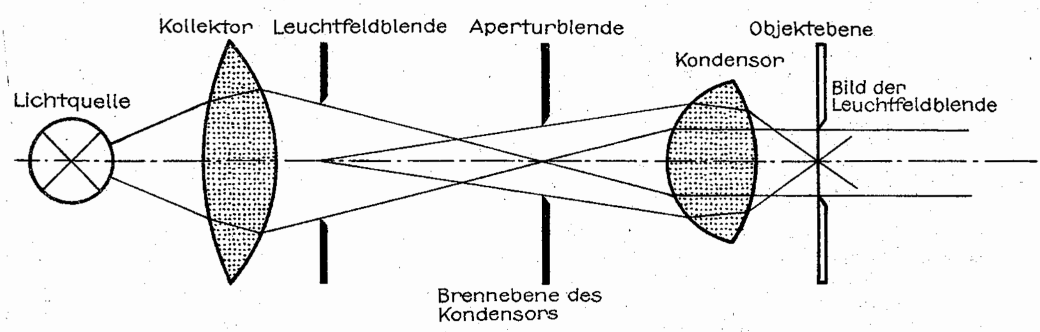


Abb. 7.5.1–9 Köhlersches Beleuchtungsprinzip

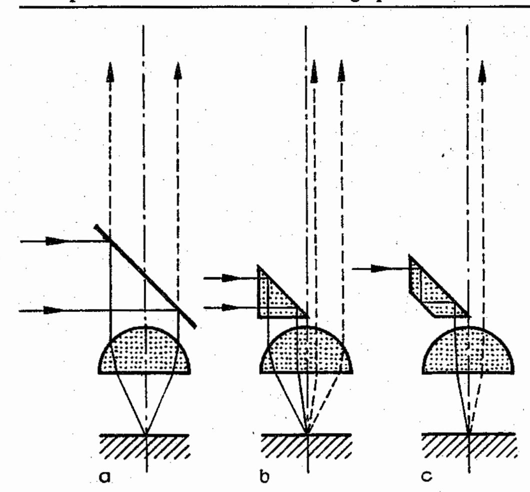


Abb. 7.5.1–10 Auflichtbeleuchtungssysteme, (a) Planglas, (b) Prisma, (c) Berek-Prisma

des Hauptmaximums geschwächt, sondern zusätzlich die Phase des direkten Lichtbündels gegenüber den Beugungsnebenmaxima um $\lambda/4$ verschoben. Die Phasenverschiebung wird mit einer planparallelen Glasplatte (*Phasenplatte*) erzeugt, in die eine ringförmige Vertiefung geätzt wurde. Die Objektbeleuchtung erfolgt über eine ringförmige Aperturblende, deren Bild mit der ringförmigen Phasenplatte übereinstimmt. Für Meßzwecke wird der Phasenkontrast selten verwandt.

Polarisationsmikroskope dienen zur Beobachtung und Vermessung anisotroper mikroskopischer Präparate mit linear polarisiertem Licht. Sie enthalten sowohl im Beleuchtungssystem als auch hinter dem Objektiv Polarisatoren. Polarisationsmikroskope werden u. a. bei der Untersuchung von Metalloberflächen und Rohprodukten für die Glas- und Keramikindustrie, in der technischen Gesteinskunde und in der Faserstoffmikroskopie eingesetzt.

Fluoreszenzmikroskope werden zur Beobachtung von Mikroskoppräparaten angewendet, die unter der Einwirkung von UV-Strahlung fluoreszieren. Derartige Präparate sind z. B. in der medizinischen und biologischen Analysentechnik, Biochemie, Zellforschung, aber auch in der Mineralogie und Geologie Untersuchungsgegenstand. Nichtfluoreszierende Objekte können mit fluoreszierenden Farbstoffen angefärbt werden. Mit stark verdünnten Fluoreszenzlösungen können z. B. lebende Zellen untersucht werden. Die Fluoreszenzausbeute liegt bei ≈ 1 bis 0,1 % der eingestrahlten Intensität. Fluoreszenzmikroskope sind meist mit Quecksilber-Höchstdrucklampen als UV-Quelle ausgerüstet. Die Beleuchtungssysteme sind für Durchlicht und Auflicht eingerichtet. Geeignete Anregungs- bzw. Sperrfilter gestatten eine Variation der Anregungs- bzw. Beobachtungswellenlänge.

7.5.1.4. Fernrohre sind optische Instrumente für die Beobachtung und Vermessung entfernter Objekte. Neben dem visuellen Gebrauch werden sie speziell in der Astronomie für die fotografische Registrierung bzw. für fotometrische oder spektralanalytische Untersuchungen eingesetzt. Fernrohre bestehen aus mindestens zwei optischen abbildenden Systemen: dem Fernrohrobjektiv und dem Okular. Bei der Beobachtung sehr weit entfernter Objekte sind beide Systeme so angeordnet, daß der bildseitige Brennpunkt des Fernrohrobjektives mit dem objektseitigen Brennpunkt des Okulars zusammenfällt. Dadurch wird ein paralleles Lichtbündel mit bestimmtem Querschnitt in ein paralleles Lichtbündel mit verändertem (in der Regel verkleinerten) Querschnitt transformiert (*teleskopischer* oder *afokaler Strahlengang*). Das Objektiv ist stets ein sammelndes optisches System, das Okular besteht aus sammelnden oder zerstreuenden optischen Bauelementen. Je nach Art der abbildenden optischen Systeme unterscheidet man zwischen Linsenfernrohren (Refraktoren) und Spiegelfernrohren (Reflektoren oder Spiegelteleskope).

Kenngrößen von Fernrohren. Wichtige optische Kenngrößen von Fernrohren sind Vergrößerung, Gesichtsfeldgröße, Lichtstärke, Bildleistung bzw. Auflösungsvermögen und Korrekturzustand der optischen Systeme. Über Vergrößerung und Gesichtsfeld werden bei den speziellen Typen Aussagen gemacht.

Lichtstärke. Bei der Beobachtung *flächenhafter Objekte* mit dem Fernrohr ist die Helligkeit (Beleuchtungsstärke E) des Netzhautbildes proportional dem Quadrat des Objektivdurchmessers D^2 (~ aufgenommener Lichtstrom) und umgekehrt proportional dem Quadrat der Vergrößerung Γ_F^2 (~ Zunahme des Netzhautbildes): $E \sim D^2/\Gamma_F^2$. Die Eintrittspupille EP fällt hierbei mit der Objektivfassung zusammen. Da der Durchmesser der Austrittspupille $D_{AP} = D/\Gamma_F$ beträgt, nimmt die Helligkeit auch proportional dem Quadrat der Austrittspupille D_{AP}^2 zu, sofern D_{AP} nicht größer als die Augenpupille des Beobachters ist. Sie ist unabhängig von der Fernrohrvergrößerung Γ_F. Die Bildhelligkeit von *Objekten mit sehr kleiner Winkelausdehnung* ($< 1'$ [z. B. Fixsterne]) nimmt mit dem Verhältnis von Objektivfläche zu Fläche der Augenpupille zu, d. h. mit dem Quadrat der Vergrößerung Γ_F^2. Die Bildhelligkeit wird vermindert durch Reflexionen an den Übergängen Luft–Glas bzw. umgekehrt (4 %/Fläche) und in geringem Maße durch Absorption im Glas. Die Reflexionsverluste können durch dielektrische Entspiegelungsschichten vermindert werden. Beim Dämmerungssehen sind Objekte umso deutlicher sichtbar, je höher die Fernrohrvergrößerung bei sonst gleichen Bedingungen ist.

Bildleistung. Die Bildleistung eines Fernrohres ist das Verhältnis von Auflösungsvermögen des Auges mit Fernrohr zum Auflösungsvermögen des freien Auges. Sie ist im Grenzfall gleich der Fernrohrvergrößerung Γ_F. Bei realen Fernrohren kann sie durch optische Abbildungsfehler und durch das Auflösungsvermögen des Fernrohres vermindert sein. Das Auflösungsvermögen wird durch Beugung an dem Objektivdurchmesser D begrenzt. Bei Abbildung eines sehr weit entfernten Objektes (Fixstern) entsteht in der Zwischenbildebene ein Beugungsscheibchen, dessen Radius (≡ Radius des 1. dunklen Beugungsringes) $r' = 1{,}22 f_{Obj} \lambda/D$ beträgt (f_{Obj} = Brennweite des Objektives), und der im Zwischenbild die minimal auflösbare Strecke darstellt. Sie entspricht einer objektseitigen Winkelauflösung $\sigma_{min} \approx r'/f_{Obj} = 1{,}22\,\lambda/D$. Die Vergrößerung des Fernrohres soll so hoch sein, daß das Beugungsscheibchen vom Auge bzw. von der fotografischen Schicht gerade noch aufgelöst wird (*Förderliche Vergrößerung*). Bei einer Winkelauflösung des Auges von 1' ergibt sich: förderliche Vergrößerung ≈ 1/2 Objektivdurchmesser in mm. Bei Hochleistungsfernrohren wird angestrebt, die optischen Systeme so gut zu korrigieren, daß die Beeinflussung der Beugungsfigur durch Abbildungsfehler vernachlässigt werden kann. Zur Kennzeichnung eines Fernrohres wird die Vergrößerung und der Durchmesser der

Abb. 7.5.1–11 Fernrohrtypen,
(a) Astronomisches,
(b) Holländisches und (c) Erdfernrohr

Eintrittspupille EP des Objektives in der Form $\Gamma_F \times D_{EP}$ angegeben. Beispielsweise hat ein Feldstecher 8 × 32 eine achtfache Vergrößerung und einen Eintrittspupillendurchmesser von 32 mm. Seine Winkelauflösung beträgt $\sigma_{min} \approx 4''$. Ein 2-m-Spiegelteleskop kann zwei Fixsterne noch bei einem Winkelabstand von $\sigma_{min} \approx 0{,}05''$ trennen.

Wichtigste Fernrohrtypen. *Astronomisches Fernrohr* (Keplersches Fernrohr). Ein unendlich fernes Objekt wird durch das sammelnde Objektiv in der gemeinsamen Brennebene von Objektiv und Okular abgebildet (Abb. 7.5.1–11a). Das sammelnde Okular wirkt wie eine Lupe in Normalvergrößerung. Es bildet das relle Zwischenbild im „Unendlichen" ab. Das erzeugte Bild ist höhen- und seitenvertauscht. Das *Gesichtsfeld* wird durch eine in der Zwischenbildebene angebrachte Feldblende scharf begrenzt. Die Fernrohrvergrößerung $\Gamma_F = \tan\sigma'/\tan\sigma$ beträgt $\Gamma_F = f'_{Obj}/f'_{Ok} = D_{EP}/D_{AP}$ (D_{EP} = Durchmesser der Eintrittspupille, D_{AP} = Durchmesser der Austrittspupille). Die Baulänge l des astronomischen Fernrohres setzt sich aus Objektivbrennweite f'_{Obj} und Okularbrennweite f'_{Ok} zusammen: $l = f'_{Obj} + f'_{Ok}$.

Holländisches Fernrohr (Galileisches Fernrohr). Es besteht aus einem sammelnden Objektiv und einem zerstreuenden Okular (Abb. 7.5.1–11b), deren Brennpunkte zusammenfallen. Das Strahlenbündel wird durch das Okular in ein paralleles Bündel überführt, bevor ein reelles Zwischenbild entsteht. Das Bild ist höhen- und seitenrichtig. Die Vergrößerung beträgt wie beim astronomischen Fernrohr $\Gamma_F = f'_{Obj}/f'_{Ok}$, seine Baulänge l ergibt sich als Differenz der beiden Brennweiten $l = f'_{Obj} - f'_{Ok}$. Das Gesichtsfeld wird nicht scharf begrenzt, da die freie Objektivöffnung als Feldblende wirkt. Holländische Fernrohre werden in der Regel als *Doppelfernrohre* ausgeführt. Ihr Prinzip wird in den Theatergläsern genutzt.

Erdfernrohre (*Terrestrische Fernrohre*) sind vom Typ des astronomischen Fernrohres, bei dem durch zusätzliche optische Bauelemente (Umkehrlinse, -prismen) eine Bildumkehr erreicht wird (*Erdfernrohre mit Umkehrlinse*). Durch eine Umkehrlinse wird von dem reellen Zwischenbild des astronomischen Fernrohres ein gleich großes, aber höhen- und seitenvertauschtes Zwischenbild erzeugt, das mit dem Okular beobachtet wird (Abb. 7.5.1–11c). In den Ebenen der reellen Zwischenbilder können Meßmarken oder Fadenkreuze angebracht werden. Die Baulänge des Fernrohres ist um den vierfachen Betrag der Brennweite der Umkehrlinse verlängert. Erdfernrohre mit Umkehrlinsen werden als Zielfernrohre eingesetzt (Typische Vergrößerungen: 1,5 bis 6×).

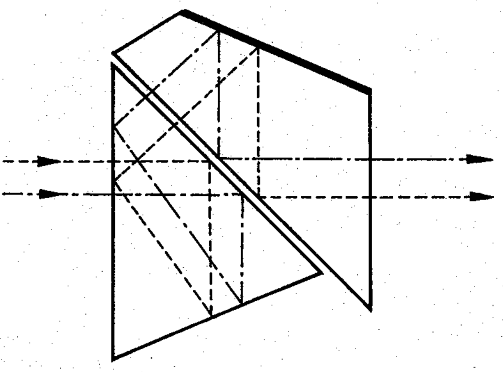


Abb. 7.5.1–13 Schmidt-Umkehrprisma

Prismenfernrohre (ABBE, PORRO). In den Prismenfernrohren werden ausschließlich totalreflektierende Prismen für die

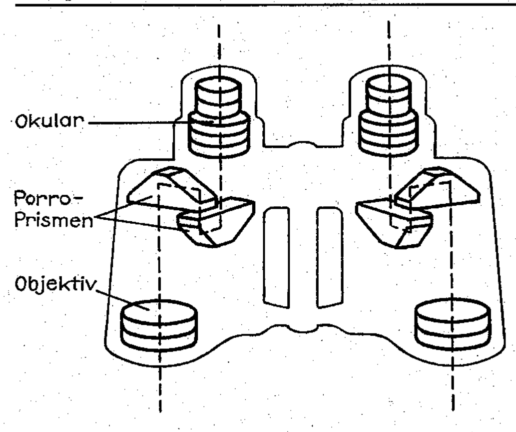
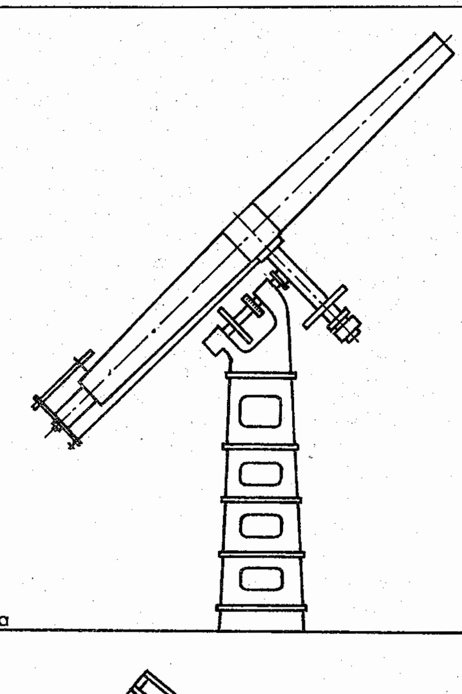
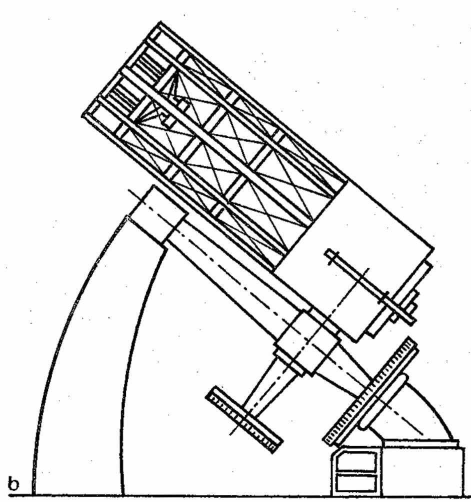


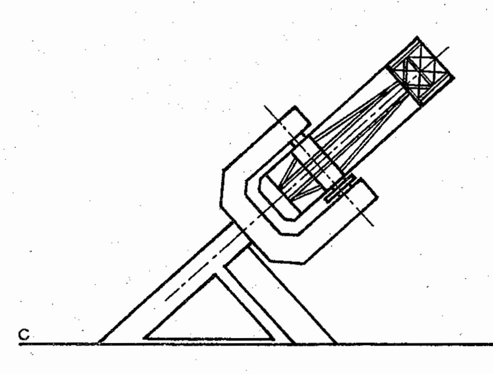
Abb. 7.5.1–12 Prismenfeldstecher

Bildumkehr eingesetzt. Weit verbreitet ist der *Porro-Prismensatz*, der von ABBE in den Fernrohrbau eingeführt worden ist. Er besteht aus vier einfachen rechtwinkligen Prismen, die durch geeignete Kombination das Bild zweimal um 90° drehen (Abb. 7.5.1–12). Das System ist geradsichtig, wobei jedoch zwischen ein- und austretendem Strahl eine Parallelversetzung auftritt. Durch Anwendung von *Schmidt-Umkehrprismen* (bestehend aus einem Dreikant- und einem Halbpentaprisma) (Abb. 7.5.1–13) kann eine Bildumkehr bei gerader Sicht ohne Parallelverschiebung erreicht werden. Prismenfernrohre werden meist als Doppelfernrohre (*binokulare Fernrohre*) gebaut. Prismenfernrohre für den Handgebrauch bezeichnet man auch als *Prismenfeldstecher*. Die maximale Vergrößerung, bei der noch eine ruhige Beobachtung mit einem Handfernrohr möglich ist, beträgt $\Gamma_F \approx 10$. Bei einem Prismenfeldstecher können die Objektive wegen der Parallelversetzung der Strahlen in den Porro-Prismen weiter auseinanderstehen als die Okulare. Ein Beobachter blickt daher mit scheinbar vergrößertem

Abb. 7.5.1–14 Astronomische Reflektoren: Strahlengänge in Spiegelteleskopen, (a) Spiegelteleskop nach NEWTON, (b) Spiegelteleskop nach CASSEGRAIN, (c) Spiegelteleskop nach COUDÉ, (d) Spiegelteleskop mit Schmidt-Kamera







Augenabstand. Deshalb nimmt der räumliche Seheindruck zu und das Bild wird plastischer. Dieser Effekt wird besonders bei *Scherenfernrohren* ausgenutzt, die einen sehr großen Objektivabstand aufweisen.

Spezielle Fernrohre für wissenschaftlichen und meßtechnischen Einsatz. *Refraktoren und Spiegelteleskope* (*Reflektoren*). Große astronomische Linsenfernrohre (Refraktoren) wurden bis zu einem Objektivdurchmesser von 1 m gebaut. Er ist begrenzt durch die große Eigenmasse der Objektivlinse, die zu Verformungen und damit zu Bildfehlern führt, da die Linse nur am Rand gefaßt werden kann. Das größte gebaute Linsenfernrohr ist der Yerkes-Refraktor in Williamsbay, USA, der einen Objektivdurchmesser $D = 1{,}02$ m und eine Brennweite $f'_{\text{Obj}} = 19{,}79$ m hat. Günstigere Bedingungen liegen bei den Spiegelteleskopen vor, deren Objektive aus großen parabolischen aluminiumbeschichteten Quarz-Hohlspiegeln bestehen, welche an der Rückseite wabenförmige Aussparungen haben und von dort entlastet werden können. Bei den großen Spiegelteleskopen befinden sich die Meßgeräte entweder im Rohrkörper im Brennpunkt des Hohlspiegels (Primärfokus) oder außerhalb des Rohres. Die Strahlung wird mittels eines kleinen Planspiegels nach außen gelenkt (*Newton-System*) (Abb. 7.5.1–14a). Durch einen konvexen Sekundärspiegel kann sie auch durch eine Öffnung in der optischen Achse des Hauptspiegels nach außen geführt werden, wobei die Brennweite vergrößert wird (*Cassegrain-System*, Abb. 7.5.1–14b). Eine Abwandlung dieses Beobachtungssystems besteht darin, daß die Strahlung über eine raumfeste Achse aus dem Teleskop heraus und in einem anderen Meßraum z. B. für spektroskopische Untersuchungen geleitet werden kann (*Coudé-System*, Abb. 7.5.1–14c). Von Bedeutung ist ferner die *Schmidt-Kamera* (Abb. 7.5.1–14d) für die fotografische Registrierung großer Gesichtsfelder. Sie stellt eine Kombination von Hohlspiegel und spezieller Korrekturplatte (*Schmidt-Platte*) zur Beseitigung der sphärischen Aberration dar. Die größten Teleskope sind in Selentschuk, UdSSR ($D = 6$ m) und auf dem Mount Palomar, USA ($D = 5{,}08$ m [Hale-Teleskop]) montiert.

Montierung (Aufstellung) von astronomischen Fernrohren erfolgt so, daß jeder Punkt des Himmels eingestellt werden kann. Dazu hat das Fernrohr zwei Drehachsen (parallel zur optischen Achse und senkrecht dazu). Je nach Lage dieser Achsen im Raum unterscheidet man zwischen *azimutaler* (nur Winkelmeßinstrumente) und der am meisten angewendeten *parallaktischen Montierung*, bei der die zur optischen Achse parallele Drehachse (Stundenachse) in Richtung der Erdachse zeigt, während die zweite Achse in der Äquatorebene liegt. Durch Drehung um die Stundenachse kann bei parallaktischer Montierung das Fernrohr einem Fixstern bei der täglichen Bewegung der Sterne nachgeführt werden. Dazu muß das Rohr mit einer Masse von einigen Tonnen mit einer Genauigkeit von $\approx 1''$ eingestellt werden. Die wichtigsten Typen parallaktischer Montierungen sind in Abb. 7.5.1–15 zusammengestellt.

Meßfernrohre sind astronomische Fernrohre, die in der Gesichtsfeldebene feste oder bewegliche Meßmarken aufweisen. Sie werden für Winkel- und Tiefenmessungen eingesetzt (Abb. 7.5.1–16).

Ablesefernrohre sind Meßfernrohre, die zur Ablesung von

Abb. 7.5.1–15 Typen parallaktischer Montierungen, (a) deutsche Montierung, (b) englische Achsenmontierung, (c) Gabelmontierung

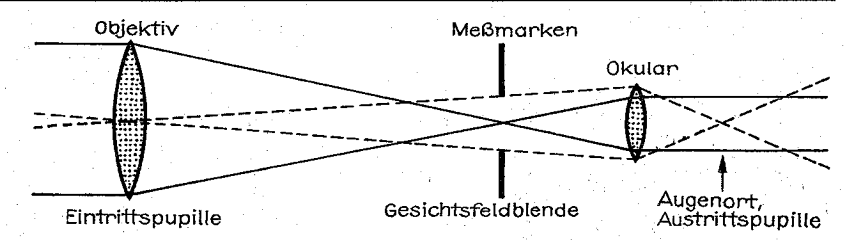


Abb. 7.5.1–16 Meßfernrohr

Marken und Teilungen in endlicher Entfernung benutzt werden.

Richtungsprüffernrohre dienen zur Messung der Geradlinigkeit und der Winkelabweichung z.B. von langen Führungen von Werkzeugmaschinen. Sie bestehen aus einer Kombination von Kollimator und Meßfernrohr.

Kollimatoren werden zur Erzeugung einer im Unendlichen erscheinenden Zielmarke eingesetzt, um ein Fernziel zu ersetzen. Die Marke befindet sich in der Brennebene des Kollimatorobjektives und wird durch eine Lampe beleuchtet. Um die Divergenz des Strahlenbündels klein zu halten (Öffnungswinkel = σ) muß wegen $\sigma \approx \tan \sigma = y/f'$ das Kollimatorobjektiv eine große Brennweite f' aufweisen und gut korrigiert sein.

Im *Autokollimationsfernrohr* sind Kollimator und Meßfernrohr vereinigt (Abb. 7.5.1–17). Die beleuchtete Meßmarke in der Objektivebene dient gleichzeitig als Zielmarke (Gaußsches Okular), die nach Reflexion an einer spiegelnden Planfläche in sich selbst abgebildet wird (Autokollimation). Voraussetzung dafür ist, daß die Planfläche genau senkrecht zur optischen Achse steht. Winkelabweichungen werden aus der Versetzung von Meßmarke und ihrem Spiegelbild gemessen. Da der Strahl bei Neigung der Planfläche um den doppelten Neigungswinkel reflektiert wird, ist das Autokollimationsfernrohr doppelt so empfindlich wie ein Richtungsprüffernrohr.

Diese passiven Richtungs- und Fluchtungsmeßverfahren haben in den letzten Jahren durch Anwendung von *Laserfluchtungsverfahren* an Bedeutung verloren. Der als Fluchtungslinie dienende rote He-Ne-Laserstrahl wird hinsichtlich seiner Lage mit einem opto-elektronischen Quadrantenempfänger vermessen.

7.5.1.5. Optische Kameras. Grundprinzip. Optische Kameras sind Geräte, die mittels eines *Fotoobjektives* von einem ebenen oder räumlichen Objekt ein reelles optisches Bild auf einer lichtempfindlichen Schicht erzeugen (Abb. 7.5.1–18).

Fotografische Kameras werden u.a. eingesetzt für die Bildfotografie, die Herstellung von Reproduktionen und Mikrofilmen, die industrielle Fotogrammetrie, die Aufnahme von Luftbildern aus Flugzeugen oder Satelliten zur topografischen Kartierung und die Erzeugung kleinster Strukturen in der Halbleitertechnologie über den fotolithografischen Prozeß. Als lichtempfindliche Materialien dienen meist gehärtete Emulsionen auf Silberhalogenid-Basis. Das durch die Belichtung erzeugte latente Bild wird erst durch einen besonderen chemisch-physikalischen Entwicklungsprozeß sichtbar. Für spezielle Zwecke werden für die fotografische Aufzeichnung fotoempfindliche Lacke, thermoplastische Filme und spezielle elektrofotografische Medien eingesetzt. In den beiden letzten Fällen wird ein latentes elektrisches Ladungsbild fixiert. Für die Aufnahme von bewegten Bildern werden neben den klassischen *Filmkameras* heute immer mehr *Video-Kameras* verwendet, bei denen das reelle

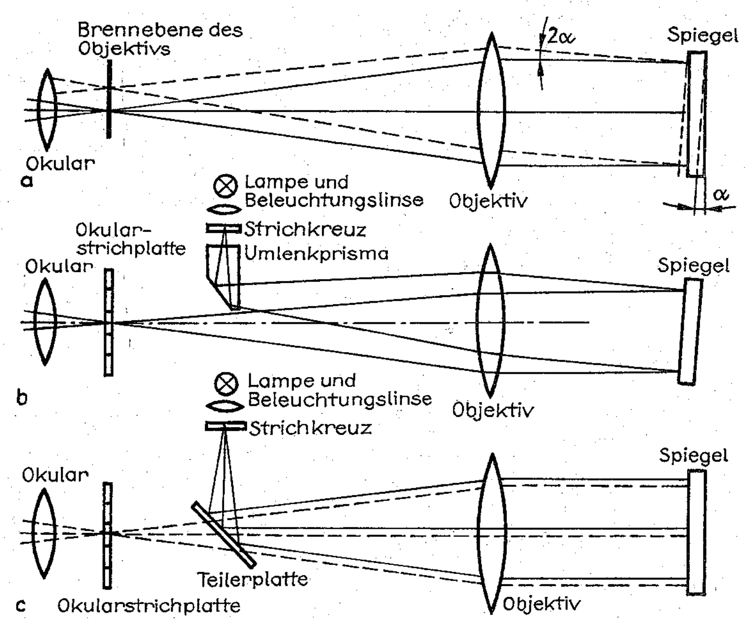


Abb. 7.5.1–17 Typen von Autokollimationsfernrohren

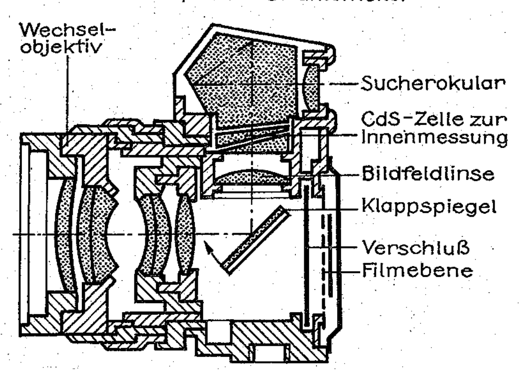


Abb. 7.5.1–18 Fotografische Kamera

optische Bild auf die Fotokatode einer Fernsehaufnahmeröhre (z. B. Vidikon) oder auf einen Halbleiter-Matrixempfänger (CCD-Matrix) abgebildet und durch den äußeren bzw. inneren fotoelektrischen Effekt in codierte elektrische Signale umgewandelt wird. Die Bilder können entweder direkt oder nach Zwischenspeicherung auf einem magnetischen Träger (z. B. in einem Video-Rekorder) mit einem Fernsehmonitor sichtbar gemacht werden. Dadurch wird der aufwendige Entwicklungsprozeß umgangen. Eine *fotografische Kamera* besteht im wesentlichen aus dem Objektiv mit Aperturblende, der Film- oder Plattenaufnahme, dem Filmtransport, dem Verschluß, der eine dosierte Belichtung ermöglicht und einem Suchersystem für die Auswahl des Objektes. Meßkameras, die in Flugzeugen bzw. Satelliten untergebracht sind, benötigen zusätzlich eine durch ihre große Relativgeschwindigkeit zur Erdoberfläche bedingte Kompensation der Bildwanderung in der Filmebene. Sie erfolgt, indem der Bildanlegerahmen während der Belichtung entgegen der Flugrichtung bewegt wird.

Fotografische Abbildung. *Abbildungsmaßstab.* Bei *Normalaufnahmen* liegt der Abbildungsmaßstab $\beta = |y'|/y$ im Wertebereich von $0 < \beta < 0{,}1$ und vergrößert sich bei *Nahaufnahmen* auf $0{,}1 \leqq \beta \leqq 1$. Der bei Nahaufnahmen erforderliche größere Abstand Objektiv–Filmebene wird durch zusätzliche Zwischenringe eingestellt.

Scharfeinstellung. Objektpunkte, deren Bildpunkte nicht in der Filmebene liegen, erzeugen dort einen Zerstreuungskreis. In der Bildfotografie gilt ein Objekt dann als „scharf" abgebildet, wenn der Zerstreuungskreis so klein ist, daß er auf dem nachvergrößerten Bild vom Auge nicht mehr aufgelöst werden kann. Für eine Kleinbildaufnahme (Negativ 25 mm × 36 mm) darf deshalb der Durchmesser des Zerstreuungskreises nicht größer als ≈ 30 µm sein, um bei einer 2,5fachen Nachvergrößerung (Positiv 60 × 90 mm²) vom Auge nicht aufgelöst zu werden. Die Scharfeinstellung des Bildes einer bestimmten Objektebene (Einstellebene) erfolgt durch Verschieben des Objektives relativ zur Filmebene.

Schärfen- und Abbildungstiefe. Alle realen Objekte haben eine räumliche Ausdehnung. Es ist daher wichtig, den Bereich des Objektraumes zu kennen, für den die zugeordneten Zerstreuungskreise in der Filmebene unter einem bestimmten Wert liegen (Abb. 7.5.1–19). Dazu werden zwei axiale Objektpunkte vor und hinter der Einstellebene betrachtet. Ihre Bilder erzeugen in der Filmebene Zerstreuungskreise. Der Bereich des Objektraumes, in dem die Zerstreuungskreise unter einem bestimmten Wert liegen, heißt *Schärfentiefe*, der zugeordnete (konjugierte) Bereich im Bildraum *Abbildungstiefe*. Mit der Entfernung der Einstellebene $a_0 \approx f/2r \cdot f/B$ ($B = f/D$ = Blendenzahl, f = Brennweite, r = Radius des Zerstreuungskreises), bei der die Schärfentiefe bis unendlich reicht, lassen sich die Grenzen der Schärfentiefe angeben:

$a_e \approx a_0/(m-1)$ und $a_r \approx a_0/(m+1)$ mit $m = a_0/a$.

Für $m = 1$ befindet sich das Objektiv in der Naheinstellung für „Unendlich". Durch Abblenden (Vergrößerung von B) wird die Schärfentiefe vergrößert.

Lichtstärke (*Öffnungsverhältnis*). Die Lichtstärke eines Fotoobjektives wird durch seine Apertur bestimmt, die bei auf unendlich eingestellten Objektiven durch das *Öffnungsverhältnis* $K = D/f = 1:B$ gegeben ist. Dabei ist D der Durchmesser der Eintrittspupille und $B = f/D$ die *Blenden- oder Öffnungszahl*. Die Beleuchtungsstärke in der Filmebene ist proportional dem Quadrat des Öffnungsverhältnisses bzw. umgekehrt proportional dem Quadrat der Blendenzahl. Bei den meisten Fotoobjektiven befindet sich die Aperturblende innerhalb des optischen Systems und die Eintritts- bzw. Austrittspupille fallen mit den Hauptebenen zusammen. Zur Anpassung der Beleuchtungsstärke an die Filmempfindlichkeit und zur Änderung der Schärfentiefe läßt sich der Durchmesser der als Irisblende ausgebildeten Aperturblende verändern. Die übliche Blendenzahlreihe (B = 0,5; 0,7; 1; 1,4; 2; 2,8; 4; 5,6; 8; 11; 16; 22) ist so aufgebaut, daß jede folgende Stufe einer Halbierung der Beleuchtungsstärke entspricht. Die maximale Lichtstärke eines Fotoobjektives ist durch den theoretisch größten Wert des Öffnungsverhältnisses K = 1 : 0,5 gegeben. Objektive mit größeren Öffnungsverhältnissen können wegen der notwendigen Einhaltung der Sinus-Bedingung prinzipiell nicht mehr korrigiert werden.

Auflösungsvermögen. Bei konventionellen Fotoobjektiven,

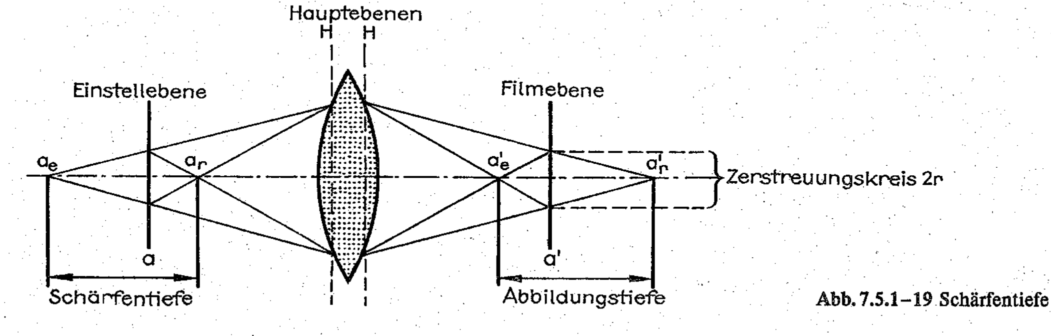


Abb. 7.5.1–19 Schärfentiefe

die für die Abbildung räumlicher Objekte konstruiert sind, wird bei vorgegebener Schärfentiefe das Auflösungsvermögen durch die Größe des geometrisch-optischen Zerstreuungskreises in der Filmebene bestimmt (Projektive Unschärfe). Spezialobjektive mit sehr hoher Auflösung haben nur einen sehr kleinen Schärfentiefenbereich. Ihr Auflösungsvermögen wird durch die Beugung an der Objektivöffnung begrenzt. Der kleinste noch auflösbare Bildpunktabstand d_{min} beträgt $d_{min} = 1{,}22\,\lambda B$. Aus meßtechnischen Gründen gibt man die Anzahl der in der Bildebene noch trennbaren Linien eines Gitters je Millimeter ($\approx 1/d_{min}$) an. Da die Beugungsfigur in der Bildebene eine räumliche Ausdehnung hat, existiert auch eine wellenoptische Abbildungstiefe $t_b = \pm 2\lambda B^2$. Wegen der Nullstellen der beugungsbedingten Intensitätsverteilung auf der optischen Achse fallen Bilddetails aus, wenn die Oberflächenunebenheiten $t_{b\,grenz}$ des Aufnahmematerials in den Bereich des Intensitätsminimums kommen ($t_{b\,grenz} \lesssim 4\,t_b$). Beugungsbegrenzte Spezialobjektive benötigen deshalb extrem ebene Aufnahmematerialien. Durch Abblenden nimmt die geometrisch-optische Unschärfe ab und die Beugungsunschärfe zu. Bei der kritischen Blendenzahl B_{krit} sind Beugungsunschärfe und Zerstreuungskreis gleich. Sie beträgt bei Kleinbildobjektiven

$B_{krit} \approx 5{,}6$ bis 8.

Perspektive. Räumliche Objekte erscheinen dem Auge unter einer bestimmten Perspektive, die vom Verhältnis der Sehwinkel hintereinander liegender Gegenstände abhängig ist. Das Auge stellt dabei das Perspektivzentrum dar, in dem sich alle Sehstrahlen schneiden. Wird von den räumlichen Objekten eine ebene Fotografie angefertigt, so sieht das Bild nur dann natürlich (d.h. mit dem gleichen perspektivischen Eindruck) aus, wenn das Verhältnis der Sehwinkel das gleiche wie bei direkter Betrachtung ist. Der perspektivische Eindruck der Fotografie hängt vom Feldwinkel des Objektives und von der Nachvergrößerung des Filmnegatives ab. Bei zu geringer Nachvergrößerung erscheint ein tiefenverlängerter, bei zu großer Nachvergrößerung ein tiefenverkürzter perspektivischer Eindruck.

Spezielle Fotoobjektive. Fotoobjektive werden charakterisiert durch Angabe der Blendenzahl B und der Brennweite f (in mm) (z.B. Flektogon 2,8/20), hochkorrigierte Objektive für die Abbildung aus endlicher Entfernung (Fotolithografische Objektive) durch ihren Abbildungsmaßstab $1:\beta$ und ihre Apertur $\sin\sigma$ (z.B. UMVM 1:5/0,25). Die Einteilung der Fotoobjektive erfolgt nach ihrem Feldwinkel $2\,\omega$, der bei vorgegebenem Filmformat durch die Brennweite f festgelegt ist. Bei einer Bilddiagonalen $2y'$ gilt dann $\tan\omega = |y'|/f$. Die Maske in der Filmebene wirkt dabei als *Feldblende*.

Bei *Normal- oder Standardobjektiven* ist die Brennweite $f \approx$ Bilddiagonale. Standardobjektive haben daher für das Kleinbildformat ($24 \times 36\ mm^2$) eine Brennweite $f = 50$ mm und für das Mittelformat ($60 \times 60\ mm^2$) $f = 80$ mm. Ihr Feldwinkel beträgt 40° bis 55°. Sie liefern bei üblichen Bildformaten den gleichen perspektivischen Eindruck wie bei direkter Objektbetrachtung.

Objektive mit Feldwinkeln > 55° werden als *Weitwinkel-Objektive*, mit Feldwinkeln > 100° bis 180° als *Überweitwinkel-Objektive* bezeichnet. Ihre Brennweiten sind kleiner als die Bilddiagonale. Sie zeichnen sich durch eine große Schärfentiefe und eine hervorgehobene Perspektive aus. Ein wichtiges Einsatzfeld dieses Objektivtyps, der dabei absolut verzeichnungsfrei korrigiert sein muß, ist die Luftbildfotogrammetrie. Objektive mit großen Brennweiten ($f > 2{,}5\,y'$) und Feldwinkeln < 35° werden als *Teleobjektive* bezeichnet. Um die große Baulänge von Teleobjektiven zu verkürzen, bestehen sie aus einem sammelnden und einem zerstreuenden Glied, damit ihre bildseitige Hauptebene weit vor dem Objektiv liegt.

Varioobjektive (Zoom-Objektive) haben eine stetig veränderliche Brennweite, so daß sich alle drei Objektivtypen in einem vereinen lassen. Dazu werden Linsengruppen vor dem feststehenden Grundobjektiv verschoben (pankratische Systeme). Vom Korrektionszustand her sind die meisten Fotoobjektive *Anastigmate*, bei denen alle Abbildungsfehler für ein relativ großes Bildfeld korrigiert sind. In Abb. 7.5.1-20 sind die wichtigsten Grundtypen der Fotoobjektive zusammengestellt. Im letzten Jahrzehnt wurden beugungsbegrenzte Spezialobjektive für die hochauflösende Fotolithografie zur Herstellung von Präzisionsskalen, -teilkreisen, -maßstäben und Masken für hoch- und höchstintegrierte mikroelektronische Schaltkreise entwickelt. Bei Bildfeldern von 3 bis 20 mm Durchmesser gestatten sie die Fertigung von Mikrostrukturen mit Strukturbreiten von 0,8 bis 2 µm bei zulässigen Restverzeichnungen von 0,02 bis 0,4 µm, indem von einer Vorlage ein 4 bis 15fach verkleinertes Bild erzeugt wird. Ihr Auflösungsvermögen beträgt bis 1500 Linien/mm. Die Ebenheit des Aufnahmematerials muß bei wenigen Mikrometern liegen.

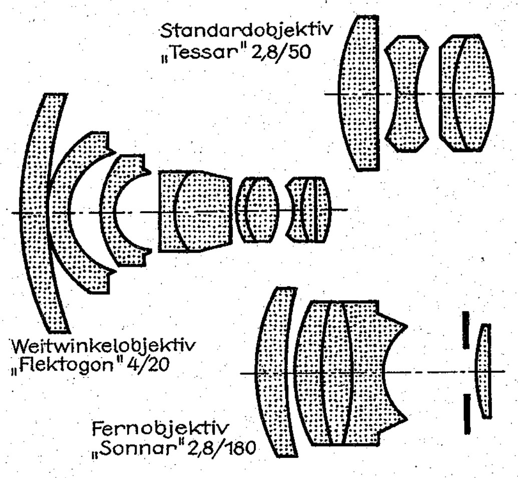


Abb. 7.5.1-20 Grundtypen von Fotoobjektiven

Hochleistungsobjektive bestehen aus Linsensystemen mit mehr als zehn zum Teil verkitteten Einzellinsen, die aus unterschiedlichen Gläsern und Kristallen gefertigt sind. Die durchschnittlichen Toleranzen für die Brechzahlen der verwendeten optischen Medien dürfen $5 \cdot 10^{-4}$ bis $1 \cdot 10^{-5}$ und die Abweichungen der individuellen optischen Achsen der Linsen von der optischen Achse des Systems 0,2′ bis 0,02′ (Dezentrierfehler) nicht übersteigen.

7.5.1.6. Bildwerfer (Diaprojektoren, Schreibprojektoren) dienen der Projektion ebener Objekte. Wesentlich ist ihr Beleuchtungssystem, welches das Objektfeld (Diapositiv, Film, Folie) mit einer hohen flächenmäßig konstanten Leuchtdichte ausleuchtet. Es besteht aus einer Projektionslampe mit großer Leuchtdichte (meist Halogenlampe), einem Reflektorspiegel und einem der Größe des Objektes angepaßten Kondensor, der die Lichtquelle in die Eintritts-

pupille des Projektionsobjektives abbildet. Das Projektionsobjektiv wirkt als umgekehrtes Fotoobjektiv mit kleiner Blendenzahl. Es bildet das Objektfeld vergrößert auf der Projektionswand ab. Zwischen der Bildweite a' und dem Abbildungsmaßstab $\beta = |y'|/y$ besteht der Zusammenhang $a' = f(1 + \beta)$.

Schreibprojektoren müssen sehr große Objektfelder (220 mm × 220 mm) abbilden. Hier werden Kondensoren aus Fresnellinsen eingesetzt.

Episkope zur Projektion undurchsichtiger Objekte werden heute nur noch selten verwendet; sie wurden durch Fernsehprojektoren abgelöst.

7.5.2. Grundtypen optisch-physikalischer Meßinstrumente

7.5.2.1. Interferometer. Grundlagen der Interferometrie. Optische Strahlungsbündel können bei Überlagerung miteinander interferieren, wenn während der Beobachtungszeit t_B zwischen ihnen feste Phasenbeziehungen vorliegen. Diese Strahlungsbündel werden *kohärent in erster Ordnung* genannt. Wenn feste Phasenbeziehungen zwischen den Teilwellen an verschiedenen Orten des Interferenzfeldes zur gleichen Zeit vorliegen, spricht man von *räumlicher Kohärenz*, wenn sie am gleichen Ort über längere Zeit bestehen von *zeitlicher Kohärenz*. Laser und thermische Lichtquellen, deren spektrale Bandbreite durch Filter eingeengt wurde und deren Strahlung beugungsbegrenzt ist, haben bei gleicher Frequenzbreite die gleiche Kohärenz erster Ordnung. Der Laser unterscheidet sich in diesem Einsatzfall nur durch wesentlich höhere Ausgangsleistung von der thermischen Lichtquelle. Bei einer Frequenzbreite $\Delta\nu_{osz}$ der Strahlungsquelle können stationäre Interferenzerscheinungen nur während einer Beobachtungszeit $t_B \ll 1/\Delta\nu_{osz}$ registriert werden. $1/\Delta\nu_{osz} = \tau_s$ ist die mittlere Strahlungszeit für einen ungestörten Wellenzug (thermische Lichtquellen $\tau_s \lesssim 10^{-8}$ s). Die in der optischen Interferometrie notwendigen längeren Beobachtungszeiten erfordern daher eine Aufteilung der Ausstrahlung in kohärente Teilwellen, die nach Durchlaufen unterschiedlicher optischer Wege überlagert werden. Sie kann durch *Teilung der Amplitude* (z. B. mittels teildurchlässiger Spiegel) oder durch *Teilung der Wellenfront* (z. B. mittels Doppelspalt oder Gitter) erfolgen. In jedem Punkt des Interferenzraumes hängt die Phasendifferenz $\Delta\varphi$ der beiden Teilwellen von der optischen Wegdifferenz Δs ab, die sich aus den unterschiedlichen Wegen der beiden Teilwellen vom Ort der Aufteilung bis zum betrachteten Aufpunkt ergibt. Die optische Wegdifferenz Δs ist proportional der geometrischen Wegdifferenz Δl: $\Delta s = n\,\Delta l$ (n = Brechzahl). Die Phasendifferenz $\Delta\varphi$ folgt daraus zu

$$\Delta\varphi = (2\pi/\lambda)\,\Delta s = (2\pi/\lambda)\,n\,\Delta l.$$

Bei der Interferenz von zwei monochromatischen Strahlungsbündeln der Intensitäten I_1 und I_2 berechnet sich dann die resultierende Intensität $I = I_1 + I_2 + 2\sqrt{I_1 I_2}\cos\Delta\varphi$.

Die begrenzte Zeit τ_s, mit der ein Wellenzug von der Strahlungsquelle ungestört abgestrahlt wird, hat seine endliche Länge $L_K = c\tau_s$ zur Folge (c = Lichtgeschwindigkeit). Gut sichtbare Interferenzstreifen treten nur auf, wenn der Gangunterschied der Teilwellen $\Delta s \ll L_K$ ist. L_K wird deshalb *Kohärenzlänge* genannt. Ein Maß für die *Sichtbarkeit* der Interferenzstreifen ist ihr Hell-Dunkel-Kontrast $K = (I_{max} - I_{min})/(I_{max} + I_{min})$.

Bei vollständiger Kohärenz und gleicher Intensität der beiden Teilbündel beträgt ihr Kontrast $K = 1$. Wenn bei gleicher Intensität $K < 1$ ist, liegt *partielle Kohärenz* vor.

Helle Interferenzstreifen, zwischen denen der Gangunterschied der Teilwellen null ist, werden als nullte Interferenzordnung bezeichnet. Links und rechts schließen sich die Interferenzstreifen 1., 2. und höherer Ordnung an. Von Ordnung zu Ordnung nimmt der Gangunterschied um einen festen Betrag zu. Die meßtechnische Ausnutzung der Interferenz erfolgt in *Interferometern*. Sie zeichnen sich durch eine hohe Meßgenauigkeit aus, da als Vergleichsgröße die Lichtwellenlänge dient. Interferometer werden für hochpräzise Längenmessungen (*Interferenzkomparatoren*), für die Vermessung von Brechzahlen bzw. Brechzahlinhomogenitäten (*Interferenzrefraktometer*), für die Winkelmessungen (z. B. *Sterninterferometer*), für Absolutmessungen von Wellenlängen mit höchster Genauigkeit und als Spektrometer mit sehr hohem Auflösungsvermögen eingesetzt. Das Prinzip der interferometrischen *Untersuchung von optischen Bauelementen* beruht darauf, daß die von dem Prüfling deformierte Wellenfront entweder mit einer idealen Wellenfront (von einem Plan- oder Konkavspiegel) oder mit sich selbst (*Verdopplungs-* oder *Shearing-Verfahren*) verglichen wird.

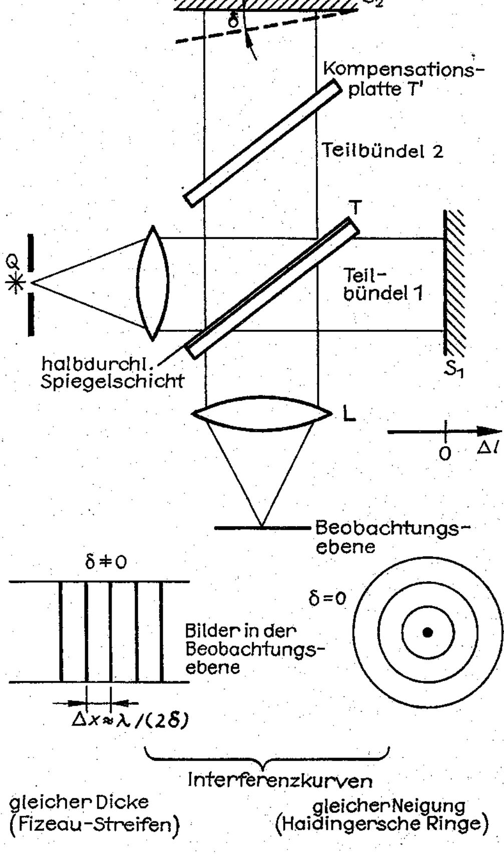


Abb. 7.5.2–1 Michelson-Interferometer

Je nach Anzahl der aufgeteilten kohärenten Bündel unterscheidet man zwischen *Zweistrahl-* und *Vielstrahlinterferometern*.

Zweistrahlinterferometer. *Michelson-Interferometer.* Das Michelson-Interferometer gehört zu den bekanntesten Zweistrahlinterferometern (Abb. 7.5.2–1). Die Strahlung einer ausgedehnten Lichtquelle Q fällt annähernd parallel auf eine, unter dem Winkel von 45° geneigte ebene Teilerplatte T, die auf der einen Seite mit einer halbdurchlässigen Spiegelschicht versehen ist. Ein Teil des Ausgangsbündels geht durch die Platte hindurch (Teilbündel 1), der andere wird reflektiert (Teilbündel 2). Beide Teilbündel sind zueinander kohärent (Amplitudenteilung). Das Bündel 1 fällt nach Transmission durch die Glasplatte auf den ebenen Spiegel S_1 und wird zur Teilerplatte T reflektiert. Das Teilbündel 2 gelangt nach Reflexion am Spiegel S_2 ebenfalls zur Teilerplatte T zurück. Damit die optischen Glaswege für beide Teilbündel gleich sind, ist zwischen Teiler T und Spiegel S_2 eine Kompensationsplatte T′ angebracht, deren Dicke genau der der Teilerplatte entspricht. Nach Vereinigung beider Teilbündel an der Teilerplatte werden sie mit der Linse L in der Beobachtungsebene des Interferometers gesammelt. Es wird zunächst angenommen, daß die optischen Wege zwischen den Spiegeln S_1 bzw. S_2 und der Teilerplatte T gleich sind. Stehen beide Spiegel S_1 und S_2 senkrecht zueinander, so registriert man in der Beobachtungsebene konzentrische helle und dunkle Kreise (*Interferenzkurven gleicher Neigung* [*Haidingersche Ringe*]). Bei Verschiebung des Spiegels 1 um die Strecke $\Delta l = \pm\lambda/4$ erfolgt im Zentrum des Ringsystems ein Wechsel von Hell nach Dunkel oder umgekehrt. Die Beobachtung der Interferenzringe geschieht visuell (direkt oder nach Vergrößerung), ihre Registrierung meist fotoelektrisch. Für Meßzwecke ist es sinnvoll, einen Spiegel um den kleinen Winkel δ aus der senkrechten Lage herauszukippen. Es entsteht dann in der Registrierebene ein Interferenzstreifensystem (*Interferenzkurven gleicher Dicke* [*Fizeau-Streifen*]), dessen Streifenabstand $\Delta x = \lambda/(2\sin\delta) \approx \lambda/(2\,\delta)$ beträgt. Wird der Spiegel S_1 in Strahlrichtung um den Betrag Δl in positiver Richtung verschoben, so wandert das Streifensystem senkrecht zur Streifenlage in negativer Richtung. Beträgt das Verhältnis der Streifenverschiebung zum Abstand zweier ruhender Streifen p, so ist die Spiegelverschiebung $\Delta l = \pm p\cdot\lambda/2$. Die Registrierung der Streifen erfolgt in der Regel fotoelektrisch (vgl. Laserinterferometer). Bei einer Meßgenauigkeit der Interferenzstreifen von 1/20 ihres Streifenabstandes ($p = 1/20$), ergibt sich eine Genauigkeit für die Ermittlung der Spiegelverschiebung von $\Delta l = \pm\lambda/40$.

Mit Hilfe des Michelson-Interferometers wurden fundamentale Untersuchungen zur Unabhängigkeit der Lichtgeschwindigkeit von der Erdbewegung durchgeführt (Michelson-Versuch, vgl. 3.1.1.3.). Das Michelson-Interferometer wird meßtechnisch u. a. eingesetzt für präzise Längenmessungen (Metervergleich, Endmaßmessung, Kontrolle an Werkzeugmaschinen), Winkelmessungen, Prüfung optischer Bauelemente und interferometrischen Formvergleich. Spezielle Ausführungsformen sind das Twyman-Green-Interferometer, die Interferenzkomparatoren, spezielle Laserinterferometer und Interferenzmikroskope sowie das Fourier-Spektrometer.

Twyman-Green-Interferometer. Es ist eine modifizierte Form des Michelson-Interferometers und wird zur interferometrischen Prüfung optischer Bauelemente (Objektivprüfung, Prüfung von Prismen sowie Prismenwinkeln und ebener Gitter) eingesetzt. Das Meßprinzip beruht auf dem Vergleich einer beim Durchgang durch den Prüfling veränderten ebenen Wellenfront (Teilbündel 1) mit einer ebenen Vergleichswellenfront (Teilbündel 2). Als Strahlungsquelle wird eine möglichst punktförmige Lichtquelle oder ein Laser verwendet. Die durch die Objektivfehler verursachten Abweichungen von der idealen Wellenfront erzeugen Interferenzstreifen, deren Auswertung die Fehler ermitteln läßt.

Interferenzkomparator. Das Michelson-Interferometer ist für präzise Relativ- und Absolutlängenmessungen gut geeignet. Eine wichtige Anwendung ist das Ausmessen von Endmaßen. Dazu wird auf dem Spiegel S_1 das Endmaß aufgesprengt und die von der Spiegeloberfläche S_1 und von der Endmaßfläche S_1' herrührenden Interferenzstreifensysteme werden miteinander verglichen. Aus der relativen Verschiebung der beiden Streifensysteme kann die Länge des Endmaßes $L = \Delta s = p\cdot\lambda/2$ ermittelt werden. In dem *Interferenzkomparator nach Köster* ist zusätzlich ein Prismenmonochromator PM vorhanden, der es gestattet, mit monochromatischem Licht zu arbeiten. Da $p \gg 1$ ist und eine gebrochene Zahl darstellt, wird zur eindeutigen Bestimmung von p meist mit drei verschiedenen Wellenlängen gearbeitet.

Laserinterferometer. Wegen der großen Kohärenzlänge frequenzstabilisierter Laser (z. B. $L_K \approx 60$ m für einen He-Ne-Laser mit einer relativen Frequenzbreite $\Delta\nu/\nu \approx 10^{-8}$) und ihrer hohen Intensität können Laserinterferometer direkt zur Längenmessung an Werkzeugmaschinen im Produktionsprozeß oder zur genauen Positionierung in der Fotolithografie eingesetzt werden. Als Meßanordnung wird meist das Michelson-Interferometer verwendet, wobei der bewegte Spiegel S_1 mit dem Meßobjekt verbunden ist. Die Zählung der Interferenzstreifen erfolgt mittels Fotodioden. Um den Einfluß von Spiegelkippungen auf das Interferenzstreifensystem auszuschließen, werden kippungsunempfindliche Spiegel (z. B. Tripelspiegel, Katzenaugenspiegel) eingesetzt (Abb. 7.5.2–2). Trotzdem kann nicht vermieden werden, daß es durch Schwingungen des mitbewegten Spiegels zu unerwünschten Nulldurchgängen an der Fotodiode kommt. Das ergibt Meßfehler, da die Fotodiode nicht unterscheiden kann, ob sich der Spiegel vor- oder rückwärts bewegt. Um eine richtungsabhängige Zählung der bewegten Interferenzstreifen zu erreichen, werden in der Nachweisebene zwei Fotodioden im Abstand der Interferenzstreifen angeordnet. Beide Dioden liefern eine je nach Bewegungsrichtung um ± 90° phasenverschobene Signalspannung, die sich zu einem elektrischen Drehfeld überlagern läßt. Die Drehrichtung liefert die Bewegungsrichtung (*Vor-* und *Rückwärtszählung*).

Mach-Zehnder-Interferometer. Es beruht wie das Michelson-Interferometer auf einer Teilung der Amplitude mittels der Teilerplatte T_1. Die beiden Teilbündel werden von den

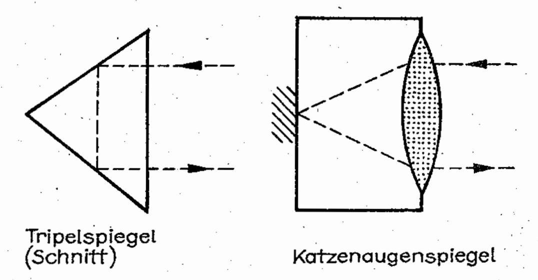


Abb. 7.5.2–2 Kippungsunempfindliche Spiegel

Spiegeln S_1 und S_2 so umgelenkt, daß sie zu einer zweiten Teilerplatte T_2 gelangen, wo sie wieder vereinigt werden (Abb. 7.5.2–3). Die Interferenzfigur wird mit einer Linse in deren Brennebene abgebildet. Je nach Justierzustand des Interferometers treten Kurven gleicher Neigung oder gleicher Dicke auf. Mach-Zehnder-Interferometer werden vorzugsweise als *Interferenzrefraktometer* zur interferometrischen Brechzahlbestimmung eingesetzt. Daraus können z.B. Strömungsprofile in Gasen oder Schlieren in Luft oder Glas ermittelt werden. Gegenüber dem ähnlichen, älteren Jamin-Interferometer hat das Mach-Zehnder-Interferometer größere getrennte Strahlenwege.

Eine Abart des Mach-Zehnder-Interferometers dient als *Shearing-Interferometer* zur interferometrischen Untersuchung optischer Baugruppen (z.B. Objektive). Shearing-Interferometer beruhen darauf, daß die beim Durchgang durch den Prüfling deformierte Wellenfront mit sich selbst verglichen wird. Dazu wird aus dem deformierten Wellenfeld (Teilstrahl 1) ein zweites gleiches Wellenfeld (Teilstrahl 2) erzeugt, das gegenüber dem ursprünglichen etwas versetzt ist. Beide Teilstrahlen werden zur Interferenz gebracht. Mit einem derartigen Interferometer kann z. B. schnell ein qualitativer Überblick über den Korrektionszustand eines Objektives gewonnen werden.

Interferenzmikroskope stellen eine Kombination von Mikroskop und Interferometer dar. Diese Mikrointerferometer gestatten an mikroskopisch kleinen Objekten aus dem Interferogramm Brechzahl, Dicke, Oberflächenformen oder Phasenänderungen bei Reflexionen am Objekt zu messen. Man unterscheidet wie bei den makroskopischen Interferometern zwei Gruppen:

1. Vergleichsstrahlengang vom Objekt unbeeinflußt: Die vom Objekt modifizierte Wellenfront wird mit einer getrennt erzeugten, idealen Wellenfront verglichen. Dazu werden sowohl Mikro-Michelson- als auch Mikro-Mach-Zehnder-Anordnungen eingesetzt (Abb. 7.5.2–4).
2. Vergleichsstrahlengang vom Objekt beeinflußt: Die vom Untersuchungsobjekt veränderte Wellenfront wird mit sich selbst verglichen, indem aus ihr eine räumlich versetzte Teilwelle mit gleicher Wellenfront erzeugt und miteinander zur Interferenz gebracht wird. Zur Strahlenteilung wird die Doppelbrechung ausgenutzt.

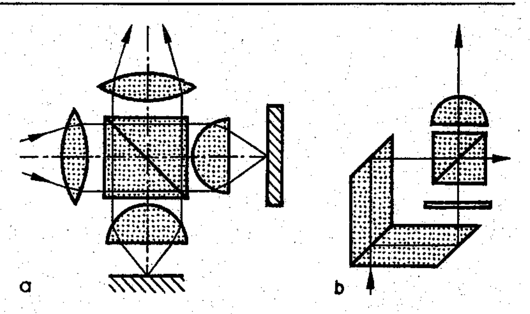


Abb. 7.5.2–4 Interferenzmikroskope, (a) Mikro-Michelson-Interferenzmikroskop, (b) Mikro-Mach-Zehnder-Interferenzmikroskop

Interferenzmikroskopie wird sowohl im Durchlicht- als auch im Auflichtverfahren angewendet.

Doppelspaltinterferometer erzeugen die beiden kohärenten Teilstrahlenbündel durch Teilung der Wellenfront. Ein sehr häufig zur Bestimmung der Brechzahl von Gasen (z.B. Grubenluft), Flüssigkeiten und Gläsern eingesetztes Doppelspaltinterferometer ist das *Rayleigh-Haber-Löwe-Interferometer* (Abb. 7.5.2–5). Mittels eines aus Eintrittsspalt und Linse bestehenden Kollimators wird ein paralleles Strahlenbündel

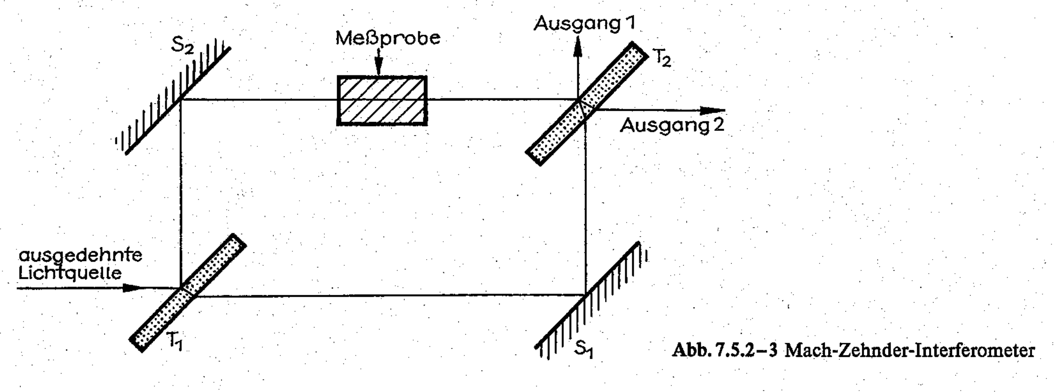


Abb. 7.5.2–3 Mach-Zehnder-Interferometer

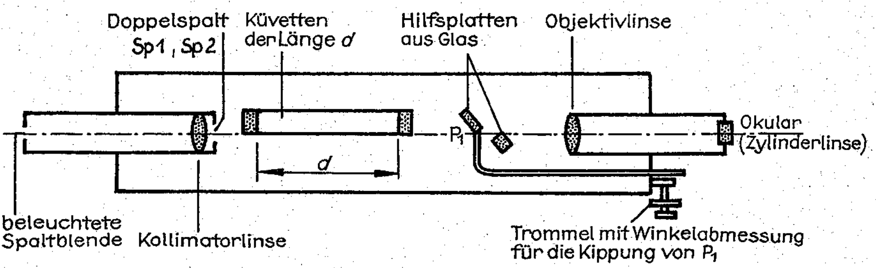


Abb. 7.5.2.–5 Rayleigh-Haber-Löwe-Interferometer

erzeugt, das mit einem Fernrohr beobachtet wird. In seiner Brennebene wird dabei der Eintrittsspalt abgebildet. Das vom Kollimator erzeugte Strahlenbündel wird durch die beiden parallelen Spalte Sp_1 und Sp_2 in zwei Teilbündel zerlegt, die miteinander interferieren. Das Spaltbild S' ist daher von Interferenzstreifen durchzogen, die parallel zum Spalt liegen. Wird in eines der beiden Teilstrahlenbündel ein Medium mit einer anderen Brechzahl gebracht, so verschiebt sich das Streifensystem. Durch Drehung einer Glasplatte P_1 in einem Strahlengang kann der erzeugte Gangunterschied kompensiert werden. Aus dem Drehwinkel wird die Brechzahl bestimmt. Bei Glasplatten von 10 mm Dicke können noch Brechzahländerungen von $\Delta n = 3 \cdot 10^{-6}$ gemessen werden.

Vielstrahlinterferometer. Die Schärfe der Interferenzlinien und damit die Meßgenauigkeit läßt sich wesentlich erhöhen, wenn das Ausgangsstrahlenbündel in viele Teilbündel zerlegt wird, die einen von Bündel zu Bündel zunehmenden festen Gangunterschied aufweisen, bevor sie wieder vereinigt werden. Eines der bekanntesten Vielstrahlinterferometer ist das *Fabry-Perot-Interferometer*. Es besteht aus zwei hochreflektierenden ebenen Spiegeln, die im Abstand l parallel zueinander stehen und einen offenen optischen Resonator bilden (Abb. 7.5.2–6). Die auf das Fabry-Perot-Interferometer unter dem Winkel Θ auftreffende monochromatische Strahlung wird durch die Vielfachreflexion zwischen den Spiegeln in viele Teilbündel aufgespalten. Der von Bündel zu Bündel fortschreitende Gangunterschied Δ der durchgehenden Strahlung beträgt $\Delta = 2nl\cos\Theta$, wobei n die Brechzahl des Mediums zwischen den Spiegeln ist. Mit einer Linsenoptik werden die austretenden parallelen Strahlenbündel in der Brennebene vereinigt. Es entstehen sehr scharfe Interferenzmaxima, die man bei durchgehendem Licht als helle konzentrische Kreise auf dunklem Untergrund beobachtet (Kurven gleicher Neigung). Maxima treten unter den Winkeln Θ auf, für die der Gangunterschied $\Delta = 2nl\cos\Theta = q\lambda$ ist (q = natürliche Zahl-Ordnung der Interferenzen). Für kleine Reflexionswinkel ($\cos\Theta \approx 1$) erhält man für die Ordnung der Interferenzen $q = 2l/\lambda$ ($n = 1$ bei Luft). Bei typischen Spiegelabständen $l = 1$ bis 500 mm erkennt man die hohe Ordnung der Interferenzen ($q = \approx 10^4$ bis 10^6). Die Interferenzen sind umso schmaler, je häufiger das Bündel zwischen den Spiegeln reflektiert wird und miteinander interferierende Teilbündel erzeugt werden. Ein Maß für die Güte des Interferometers stellt seine „Finesse" $\tilde{N}$ dar, $\tilde{N} = \pi\sqrt{R}/(1-R)$, die als effektive Zahl der miteinander interferierenden Strahlen interpretiert werden kann. R bedeutet darin den totalen Reflexionskoeffizienten des Fabry-Perot-Interferometers: $R = \sqrt{R_1 R_2}$ (R_1, R_2 = Teilreflexionsfaktoren der beiden Spiegel). Das Verhältnis von durchgehender Leistung P_t zu auffallender Leistung P_0 wird durch die *Airy-Formel* beschrieben

$$\frac{P_t}{P_0} = \frac{T^2}{(1-R)^2} \frac{1}{1 + 4R/(1-R)^2 \sin^2 \Phi/2},$$

wobei $\Phi = (2\pi/\lambda)\Delta$ die Phasendifferenz zwischen zwei Teilbündeln und $T = \sqrt{T_1 T_2}$ die totale Transmission des Fabry-Perot-Interferometers angeben (vgl. Abb. 7.5.2–6). Man erkennt, daß die Interferenzlinien umso schmaler sind, je größer die „Finesse" des Interferometers ist. Sie wird außer vom Reflexionsvermögen bei „idealen" Spiegeln von Plattenunebenheiten, Justierfehlern und Beugungsverlusten bei endlichen Bündeldurchmessern beeinflußt. Für Fabry-Perot-Spiegel ist Quarz als Trägermaterial gut geeignet. Es läßt sich für einen Spektralbereich von 200 nm bis 3,2 µm einsetzen. Plattenebenheiten von $\lambda_{grün}/50$ über einen Durchmesser von 50 bis 60 mm sind leicht erreichbar. Mit dielektrischen Verspiegelungen werden Werte von $\tilde{N} \approx 100$ erzielt. Die Spiegelplatten haben einen Keilwinkel ($\approx 0{,}5°$ bis 1°), damit keine unerwünschten Interferenzen in den Platten auftreten. Fabry-Perot-Interferometer können auch in Reflexion betrieben werden. Sie werden zur absoluten Wellenlängenmessung, zur Frequenzfilterung und als hochauflösende Spektrometer eingesetzt (vgl. 7.5.2.2.). Bei Ausführungen mit festem Plattenabstand ($l \approx 0{,}3$ bis 200 mm) erfolgt die Registrierung fotografisch. Eine Änderung der optischen Länge kann piezoelektrisch oder durch Änderung des Gasdruckes vorgenommen werden. Für genaue Messungen müssen der Luftdruck und die Temperatur (Einfluß auf Brechzahl und Plattenabstand) konstant gehalten werden.

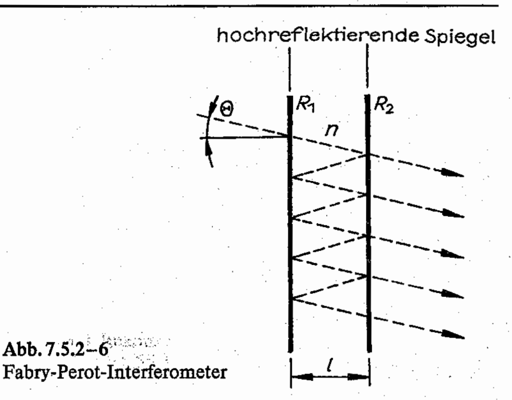


Abb. 7.5.2–6
Fabry-Perot-Interferometer

Hologramminterferometer. Holografische Aufzeichnungsverfahren gestatten die zweidimensionale Aufzeichnung von räumlichen Objekten und ihre dreidimensionale Rekonstruktion. Dazu wird das Objekt mit einem Laser beleuchtet. Das gestreute Licht (Objektwelle) wird mit einer kohärent abgeteilten Referenzwelle in der fotografischen Schicht zur Interferenz gebracht (Abb. 7.5.2–7). Dadurch wird neben der Amplitude die Phase der Objektwelle gespeichert. Durch Beleuchtung des Hologramms mit einem Laser wird in dem erzeugten Beugungsbild das Objekt rekonstruiert. Durch Kombination der Holografie mit interferometrischen Methoden können kleine räumliche Veränderungen des Objektes im Bereich von $\approx 0{,}1$ µm bis einige 100 µm gemessen werden. Die Aufnahme des holografischen Interferogramms kann in *Doppelbelichtungstechnik* erfolgen, indem das Objekt zeitlich nacheinander mit einem kurzen Laserimpuls belichtet wird, und die beiden Hologramme in der Fotoplatte übereinander registriert werden. Durch Überlagerung wird ein interferometrischer Vergleich der beiden verschobenen Objektlagen erreicht. Mit Hilfe der *Echtzeittechnik* können nur die Veränderungen am Objekt sichtbar gemacht werden. Dazu wird von dem Objekt ein Hologramm angefertigt, das nach der Entwicklung exakt an die Stelle gebracht wird, wo die Fotoplatte belichtet wurde. Bei Rekonstruktion des Objektes mittels der Referenzwelle entsteht das virtuelle Bild genau am Ort der ursprünglichen Lage des Objektes. Bei gleichzeitiger Beleuchtung des Objektes mit den Laser entstehen Interferenzen zwischen dem gestreuten Laserlicht und dem für die Rekonstruktion verwendeten an den Orten, wo Lageänderun-

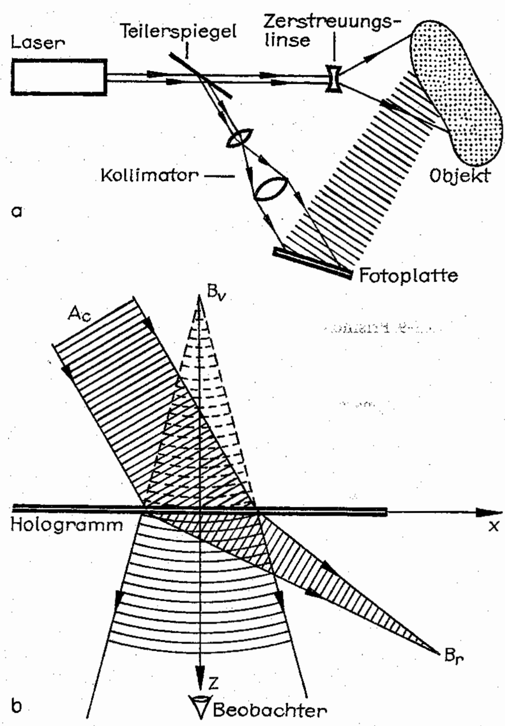


Abb. 7.5.2–7 Hologramminterferometer, (a) Grundschema der Auflichtholografie, (b) Rekonstruktion des Hologramms mit einer ebenen Referenzwelle

gen aufgetreten sind. Dieses Verfahren ist auch zur Fehlererkennung bei mikroelektronischen Schaltkreisen geeignet. Wird das Hologramm von einem einwandfreien Schaltkreis hergestellt, so treten beim Vergleich mit den Prüflingen nur dort Interferenzmuster auf, wo es zu Abweichungen von dem Prüfnormal kommt.

Weitere kohärent-optische Methoden. *Speckle-Interferometrie*. Wird eine rauhe Oberfläche mit der kohärenten Strahlung eines Lasers beleuchtet, so beobachtet man mit dem Auge eine kontrastreiche grobkörnige Intensitätsverteilung im diffus reflektierten Licht. Durch Bewegung des Auges läßt sich diese Intensitätsstruktur abtasten. Wird die Oberfläche verschoben, so wandert das Granulationsmuster mit. Diese optische Erscheinung ist typisch für Laserlicht, da wegen seiner hohen Interferenzfähigkeit Laserlichtwellen, die von unterschiedlichen Oberflächenpunkten ausgehen, miteinander interferieren können. Die Rauhigkeit der Oberflächen ist die Ursache für die statistisch verteilte, grobkörnige Intensitätsstruktur, deren mittlerer Durchmesser d auf einer Fotoplatte bei auf „Unendlich“ eingestelltem Fotoobjektiv näherungsweise durch $d \approx (\lambda/D) \cdot f = B\lambda$ gegeben ist (D = Durchmesser der Aperturblende, B = Blendenzahl, f = Brennweite). Diese Intensitätskörner werden üblicherweise *Speckles* (engl., Flecken) genannt. Das mit der Oberfläche eines Gegenstandes verbundene Specklemuster gestattet es, Verschiebungen, Verformungen oder Schwingungen des Körpers fotografisch zu vermessen. Dazu wird das Specklemuster vor und nach der Verformung registriert. Die gegenseitige Verschiebung der einzelnen Speckles auf der Fotoplatte führt bei Durchstrahlung mit parallelem monochromatischem Licht zu Interferenzen in der Brennebene einer Linse. Diese Interferenzfigur entspricht dem Beugungsbild zweier kohärenter Lichtquellen im Abstand der verschobenen beleuchteten Speckles. Gegenüber der Hologramminterferometrie ist die Specklemethode weniger anfällig gegenüber äußeren Störungen, z. B. Schwingungen. Sie ist jedoch auch weniger exakt. Speckleverfahren eignen sich gut für eine elektronische Bildverarbeitung.

Moiré-Verfahren registrieren fotografisch Veränderungen selbst sehr großer Objekte (Flugzeuge, Schiffe). Dazu wird ein Lichtgitter auf dem Objekt erzeugt, das vor und nach der Veränderung fotografiert wird. Durch die Überlagerung der verformten Lichtgitter in der fotografischen Schicht entsteht ein charakteristisches Moiré-Muster. Werden die Lichtgitter interferometrisch erzeugt, kann ihre Gitterkonstante bequem dem Objekt angepaßt werden.

Sensoren auf der Basis von Lichtleiter-Interferometern. Lichtwellenleiter bestehen aus einem Kern aus hochbrechendem Glas, der von einem Glasmantel mit niedrigerer Brechzahl umgeben ist. Bei *Einmoden-* oder *Monomode-Lichtwellenleitern* ist der Durchmesser des Kerns soweit reduziert, daß sich nur eine Lichtwelle mit einer definierten Feldverteilung ausbreiten kann. Die Kerndurchmesser betragen ≈ 3 bis 10 µm. Dadurch ist die Phase Φ längs der Ausbreitungsrichtung eindeutig definiert: $\Phi(z) = kz = 2\pi/\lambda_0 \cdot z$. Dieser Effekt wird in Monomode-Sensoren ausgenutzt. Sie beruhen darauf, daß eine Meßgröße P (Temperatur, Druck, Spannung, magnetische Feldstärke, elektrische Feldstärke) über einen geeigneten physikalischen Effekt (thermische Ausdehnung, Dehnung, Elektrostriktion, Magnetostriktion usw.) so auf die Faser einwirkt, daß die Phase Φ_M geändert wird. Zur Messung wird die Phase Φ_M mit der Phase in einem Vergleichszweig verglichen, der unabhängig von P ist. Dazu werden spezielle miniaturisierte Interferometeranordnungen in Lichtleitertechnik, wie Michelson- und Mach-Zehnder-Interferometer zur Messung schnell veränderlicher Größen, Sagnac-Interferometer für die Ermittlung von Winkelgeschwindigkeiten und -beschleunigungen und Fabry-Perot-Interferometer als optische Spektralanalysatoren und zur Wellenlängenstabilisierung von Laserdioden angewendet. Es werden Meßgenauigkeiten von $\Delta\Phi \approx 10^{-5}$ bis 10^{-6} rad erreicht. Diese Sensoren zeichnen sich durch Kompaktheit und Robustheit bei hoher Meßempfindlichkeit aus. Glasfaser-Fabry-Perot-Interferometer wurden mit Auflösungsvermögen $> 5 \cdot 10^6$ gebaut.

7.5.2.2. Spektralapparate sind Geräte, die für die Beobachtung, Registrierung bzw. Messung der wellenlängenabhängigen Transmission, Absorption, Reflexion oder Streuung optischer Strahlung eingesetzt werden. Im engeren Sinne bezeichnet man damit Geräte, die eine wellenlängenabhängige (spektrale) Zerlegung optischer Strahlung gestatten. Einfache Spektralapparate, die nur eine visuelle Beobachtung ermöglichen, werden *Spektroskope* genannt. Registrierende Spektralapparate heißen *Spektrografen*. Als *Spektrometer* werden Spektralapparate (im engeren Sinne) bzw. komplette Anordnungen (bestehend aus Strahlungsquelle, Wechselwirkungsteil Strahlung–Probe, eventuell Geräte zur spektralen Zerlegung, Nachweisanordnung) bezeichnet, die vorzugsweise für die Messung spektral abhängiger physikalischer Größen eingesetzt werden. Spektralapparate, die aus einem Wellenlängengemisch eine monochromatische Strahlung herausfiltern, nennt man *Monochromatoren*. Bei

einem *Laserspektrometer* wird die monochromatische (z. T. frequenzdurchstimmbare) Strahlung in der Laser-Strahlungsquelle selbst erzeugt, so daß unter Umständen auf eine zusätzliche spektrale Zerlegung verzichtet werden kann.

Grundlagen der Spektralmessungen. Man unterscheidet zwischen Spektrometern mit breitbandigen Strahlungsquellen, deren Emission kontinuierlich oder diskret in einem breiten Frequenzbereich erfolgt (z. B. Hochdruck-Gasentladungslampen, thermische Strahler, Linienstrahler bei Funkenanregung), die einer spektralen Zerlegung bedürfen, und Spektrometern mit monochromatischen (eventuell durchstimmbaren) Strahlungsquellen (Lasern), bei denen eine spektrale Zerlegung nur erforderlich ist, wenn bei der Wechselwirkung der Strahlung mit der Probe neue Frequenzen entstehen.

Für die spektrale Zerlegung optischer Strahlung sind eine Reihe von Methoden anwendbar.

- Die Abhängigkeit der Brechzahl optischer Materialien von der Wellenlänge (*Materialdispersion*) führt in Prismen zur Wellenlängenabhängigkeit des Ablenkwinkels.
- Durch Beugung entstehen beim Durchgang oder bei der Reflexion an einem Beugungsgitter Nebenmaxima, deren Lage wellenlängenabhängig ist (*Beugungsdispersion*).
- Durch *Interferenz* von Strahlenbündeln großer Gangunterschiede entstehen wellenlängenabhängige Interferenzmuster.

Spektralapparate auf der Basis spektraler Zerlegung optischer Strahlung lassen sich daher einteilen in: *Prismenspektralapparate*, *Gitterspektralapparate* und *Interferenzspektralapparate*.

Kenngrößen von Spektralapparaten. *Apparatefunktion*. Am Ausgang eines Spektralapparates erzeugt das auffallende Leistungsspektrum $S_w(\lambda)$ ein Signal $S(\lambda)$, das durch unvermeidliche *systematische* und *zufällige Fehler* verfälscht ist. Die systematischen Fehler des Spektralapparates werden durch seine *Apparatefunktion* $a(\lambda)$ gekennzeichnet, welche die Verzerrung des Eingangsspektrums durch die optischen und elektrischen Bauelemente beschreibt. Bei vernachlässigbaren statistischen Fehlern kann die Apparatefunktion $a(\lambda)$ dadurch gemessen werden, daß das Signal („response“) eines monochromatischen Eingangsspektrums

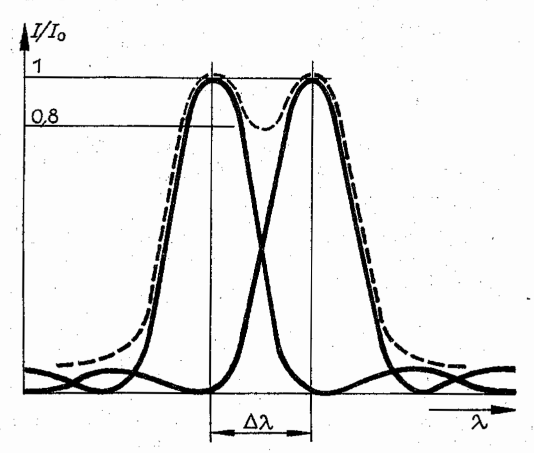


Abb. 7.5.2–8 Auflösung zweier Spektrallinien

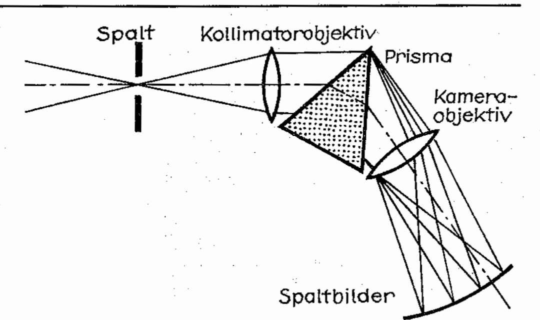


Abb. 7.5.2-9 Prismenspektralapparat

aufgezeichnet wird. Das registrierte Signal $S(\lambda)$ ist dann gleich der Apparatefunktion $a(\lambda) = S(\lambda)$. Die Halbwertsbreite der Apparatefunktion $(\Delta\lambda)_a$ entspricht näherungsweise dem kleinsten auflösbaren Linienabstand.

Auflösungsvermögen. Es wird üblicherweise festgelegt (Abb. 7.5.2–8), daß zwei eng benachbarte monochromatische Spektrallinien annähernd gleicher Intensität, die einen Wellenlängenabstand $\Delta\lambda$ haben, durch den Spektralapparat dann noch getrennt registriert, d. h. „aufgelöst“ werden, wenn die Intensität des registrierten Signals zwischen den Maxima beider Linien mindestens auf $\approx 80\,\%$ abgefallen ist (*Rayleigh-Kriterium*). Das Auflösungsvermögen des Spektralapparates beträgt $A = \lambda/\Delta\lambda$.

Dispersionsgebiet. Der ausnutzbare Wellenlängenbereich eines Spektralapparates (Meßbereich), der gleichzeitig ungestört beobachtet werden kann, wird *Dispersionsgebiet* genannt. Prismenspektralapparate haben ein großes Dispersionsgebiet, das lediglich durch Totalreflexion an der Austrittsfläche des Prismas bzw. durch einsetzende Absorption des Prismenmateriales begrenzt wird. Das Dispersionsgebiet von Gitter- und Interferenzspektralapparaten wird durch die Überlagerung verschiedener Ordnungen begrenzt. Es ist umso kleiner, je größer das Auflösungsvermögen ist. Bei hochauflösenden Geräten ist eine Vorzerlegung erforderlich.

Nachweisempfindlichkeit bzw. Lichtstärke. Das von einem Spektralapparat registrierte Signal ist durch statistische Fehler verfälscht, die ihre Ursache in der Quantennatur der *Photonen* (*Photonenrauschen*), in den Rauschprozessen der Fotoplatte bzw. der fotoelektrischen Empfänger und in einem eventuellen Streulichtuntergrund haben. Die Nachweisempfindlichkeit wird durch das Verhältnis von Signal- zu Rauschleistung des registrierten Signals bestimmt. Als untere Grenze für den Nachweis eines Signals läßt sich festlegen, daß die Signalleistung größer oder mindestens so groß wie die Rauschleistung sein muß. Die nachgewiesene Signalleistung ist umso größer, je größer die *Lichtstärke* des Spektralapparates ist.

Spektrometer mit dispergierenden Funktionselementen. *Prismenspektralapparate*. Den Grundaufbau eines Prismenspektralapparates zeigt Abb. 7.5.2–9. Die durch den schmalen Eintrittsspalt S_1 einfallende optische Strahlung wird mit der Kollimatorlinse L_1 in ein paralleles Strahlenbündel umgewandelt, welches das *Prisma* P ausleuchtet. Aufgrund der Wellenlängenabhängigkeit der Brechzahl $dn/d\lambda$ (= Materialdispersion) wird die Strahlung beim Durchgang durch das Prisma in Abhängigkeit von der Wellenlänge unterschiedlich abgelenkt. Von diesen Parallelstrahlenbündeln erzeugt die Linse L_2 mit der Brennweite f_2 in der Brenn-

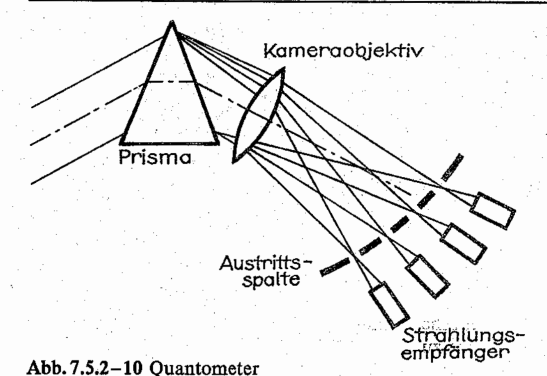


Abb. 7.5.2–10 Quantometer

ebene nebeneinanderliegende monochromatische Bilder (genauer Beugungsfiguren) des Eintrittsspaltes S_1, die sogenannten *Spektrallinien*. Sie können entweder durch ein Okular (Lupe) beobachtet oder mittels einer fotografischen Schicht bzw. eines oder mehrerer fotoelektrischer Empfänger registriert werden.

Wenn nur ein Teil des Spektrums erfaßt werden soll, wird in der Brennebene von L_2 ein *Austrittsspalt* S_2 regelbarer Breite angebracht. Spektrografen, die quantitativ die Intensität bestimmter Spektrallinien bzw. deren Intensitätsverhältnis fotoelektrisch registrieren (*Quantometer*), sind mit mehreren Austrittsspalten ausgestattet (Abb. 7.5.2–10). Geräte, die aus einem vorliegenden Leistungsspektrum nur einen bestimmten, möglichst schmalen Wellenlängenbereich ausblenden, werden als *Monochromatoren* bezeichnet.

Winkeldispersion. Bei symmetrischem Durchgang durch das Prisma (Minimum der Ablenkung) beträgt die Änderung des Ablenkwinkels $d\alpha$ mit der Wellenlängenänderung $d\lambda$:

$$\frac{d\alpha}{d\lambda} = \frac{2\sin\varphi/2}{(1-n^2\sin^2\varphi/2)^{1/2}}\,\frac{dn}{d\lambda}$$

($dn/d\lambda$ = Materialdispersion, φ = brechender Winkel des Prismas, n = Brechzahl des Prismenmateriales).

Der Abstand zweier Spaltbilder (Spektrallinien) in der Nachweisebene wird durch die *Lineardispersion* bestimmt: $dl/d\lambda = f_2\, d\alpha/d\lambda$.

Die *Lichtstärke* des Spektralapparates ist durch sein Öffnungsverhältnis D/f gegeben (D = freie Öffnung des Kollimators, f = Kollimatorbrennweite).

Das *Auflösungsvermögen* des Prismenspektralapparates wird durch die Beugung des Parallelbündels an dem Prismenquerschnitt begrenzt. Es hängt bei voller Ausleuchtung des Prismas von dessen Basisbreite $\bar{B}$ und von der Materialdispersion ab, $A = \lambda/\Delta\lambda = -\bar{B}\, dn/d\lambda$.

Es ist am größten beim symmetrischen Durchgang durch das Prisma. Das Auflösungsvermögen kann durch Hintereinanderschalten von Prismen erhöht werden.

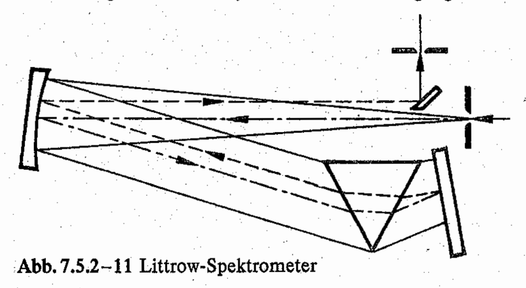


Abb. 7.5.2–11 Littrow-Spektrometer

Spezielle Prismenspektralapparate. Prismenspektrografen mit einer Fotoplatte als Empfänger sind im wesentlichen auf den ultravioletten und sichtbaren Spektralbereich beschränkt. Im UV-Bereich werden *Quarzspektrografen* mit einfachen Quarzlinsen eingesetzt. Wegen der wellenlängenabhängigen Brennweite der Abbildungsoptik muß die Fotoplatte schräg zur Einfallsebene angeordnet werden.

Da die Linsenoptiken mit Nachteilen verbunden sind, werden Spektrografen häufig mit sphärischen Spiegeloptiken ausgestattet. Eine einfache und sehr häufig angewendete Bauart stellt die *Autokollimationsanordnung* nach LITTROW (Abb. 7.5.2–11) dar. Der Kollimatorspiegel dient hier sowohl zur Parallelisierung der vom Eintrittsspalt kommenden Strahlung, als auch zur Abbildung der Spektrallinien in die Bildebene. Wegen des zweifachen Durchganges der Strahlung durch das Prisma wird die doppelte Auflösung erzielt.

Prismenmonochromatoren benötigen einen gekrümmten Austrittsspalt, weil die schräg durch das Prisma hindurchlaufenden Strahlen stärker abgelenkt werden als die im Hauptschnitt verlaufenden. Zur Erhöhung des Auflösungsvermögens und der Verminderung des Streulichtes werden *Doppelmonochromatoren* eingesetzt.

Gitterspektralapparate. Ein Gitterspektralapparat ist mit einem *Beugungsgitter* als dispergierendes Element ausgestattet. Es werden fast ausschließlich *Liniengitter* verwendet, die

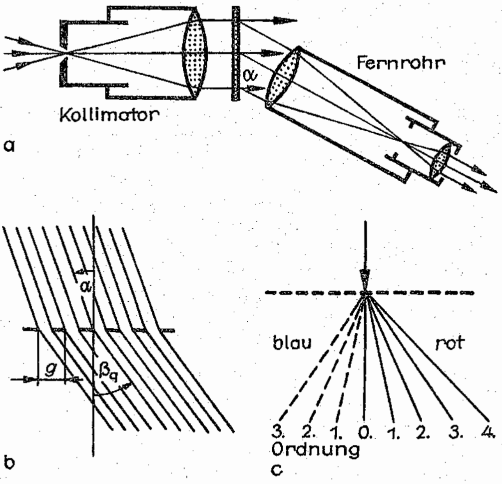


Abb. 7.5.2–12 Beugungsgitter, (a) Schema des Gitterspektroskops, (b) Strahlengang am Beugungsgitter (c) Überlagerung der Spektren unterschiedlicher Ordnung

aus äquidistanten feinen, parallelen Spalten in einem undurchsichtigen Schirm oder aus schmalen Spiegelelementen bestehen. Die Beugungsgitter sind meist eben (*Plangitter*), für spezielle Zwecke (z. B. Wegfall der zusätzlichen Optik im VUV-Spektralbereich) werden gekrümmte Gitter (Teilung parallel zur Zylinderachse) hergestellt (*Konkavgitter*). Das Beugungsgitter wird charakterisiert durch die *Gitterkonstante* g (= Abstand zwischen entsprechenden Punkten zweier benachbarter Gitterfurchen) bzw. ihrem Reziprokwert $\bar{n} = 1/g$ (= Anzahl der Linien/mm) und der *Gesamtzahl der Linien* N. Der Grundaufbau eines Gitter-

spektralapparates ist analog dem eines Prismenspektralapparates. Mittels eines Kollimators wird erreicht, daß die vom Eintrittsspalt kommende Strahlung als möglichst paralleles Strahlenbündel auf das Gitter fällt (*Frauenhofersche Beugung*). Die freie Öffnung der Kollimator- und Abbildungsoptik muß so groß sein, daß die Gitterfläche voll ausgeleuchtet wird. In der reellen Bildfläche werden die gebeugten Strahlen vereinigt, und es entstehen bei unendlich schmalem Eingangsspalt monochromatische Beugungslinien der Breite $b = 2f\delta\varphi$, wobei $\delta\varphi$ der Winkel zwischen Beugungsmaximum und 1. Minimum ist. Es gilt $\delta\varphi = \lambda/Ng = \lambda/B$ (B = Gitterbreite).

Lage der Beugungsmaxima. Wenn optische Strahlung unter dem Winkel α zur Gitternormalen einfällt (Abb. 7.5.2–12), so entstehen beiderseits der nullten Beugungsordnung Beugungsmaxima der 1., 2., 3., ..., q-ten Ordnung, die unter dem Ablenkwinkel $\varphi q = \alpha + \beta_q$ auftreten:

$$\sin\alpha + \sin\beta_q = 2\sin\left(\frac{\varphi q}{2}\right)\cos\left(\frac{\alpha - \beta_q}{2}\right) = \frac{q\lambda}{g}$$

($\beta_q < 0$, wenn der gebeugte Strahl auf der entgegengesetzten Seite der Gitternormalen liegt wie der einfallende Strahl). Bei senkrechtem Einfall ($\alpha = 0$) gilt: $\sin\varphi_q = q\lambda/g$. Daraus folgt, daß mit steigender Ordnung und mit zunehmender Wellenlänge die Beugungslinien auseinanderrücken. Rotes Licht wird stärker abgelenkt als blaues. Die *Winkeldispersion* des Gitters beträgt (α = const.): $\mathrm{d}\varphi/\mathrm{d}\lambda = q/(g\cos\beta)$. Für kleine Beugungswinkel β ($\cos\beta \approx 1$) ist die Winkeldispersion konstant $\mathrm{d}\varphi/\mathrm{d}\lambda \approx q/g$ (gleiche Abstände im Spektrum entsprechen dann gleichen Wellenlängendifferenzen). Ein derartiges Spektrum wird *Normalspektrum* genannt. Es eignet sich besonders für Wellenlängenmessungen.

Auflösungsvermögen A. Für kleine Beugungswinkel ergibt sich $A = \lambda/\Delta\lambda = qN$. Diese Gleichung gilt nur für unendlich schmale Eintrittsspalte und ideale Abbildung. Das theoretische Auflösungsvermögen nimmt mit steigender Ordnung zu.

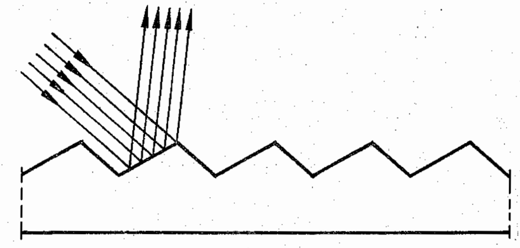


Abb. 7.5.2–13 Geblaztes Gitter

Dispersionsgebiet $\Delta\lambda_D$. Fällt polychromatische Strahlung auf das Gitter, so überlagern sich die Spektren höherer Ordnung. Der ungestört ausnutzbare Wellenlängenbereich $\Delta\lambda_D$ (Dispersionsgebiet) beträgt $\Delta\lambda_D = \lambda/q$. Er nimmt mit steigender Ordnung ab.

Spezielle Beugungsgitter, geblazte Gitter. Beugungsgitter werden meist als *Reflexionsgitter* mit einem sägezahnartigen Furchenprofil hergestellt (Abb. 7.5.2–13). Derartige Gitter konzentrieren die reflektierte Strahlung vorzugsweise in eine bestimmte Ordnung des Beugungsspektrums *(Blaze-Bereich)*. Damit wird ein Nachteil der spektralen Zerlegung mit Beugungsgittern gegenüber Prismen überwunden, der darin besteht, daß sich üblicherweise bei einem Gitter die eingestrahlte Leistung auf alle Beugungsordnungen verteilt. Geblazte Gitter heißen auch *Echelette-Gitter*.

Beugungsgitter. Optische Beugungsgitter im UV-, VIS- und NIR-Bereich werden in einem sehr aufwendigen Verfahren auf Gitterteilmaschinen mittels einer feinen Diamantspitze in dünne metallische Aluminium- oder Goldschichten von einigen Mikrometern Dicke geritzt, die auf Glas oder

Tab. 7.5.2–14 Auflösungsvermögen und Dispersionsgebiet von Spektralapparaten

	Auflösungsvermögen $A = \lambda/\Delta\lambda$	Dispersionsgebiet $\Delta\lambda_D, \Delta\nu_D$	experimentelle Werte A	$\Delta\lambda_D, \Delta\nu_D$
Prismen-Spektral-apparate	$A = \lambda/\Delta\lambda = -B^{\mathrm{d}n}/\mathrm{d}\lambda$ B = Basisbreite $\mathrm{d}n/\mathrm{d}\lambda$ = Material-dispersion	abhängig vom Prismen- bzw. Linsenmaterial	$A \lesssim 7\cdot10^3 \ldots 5\cdot10^4$ $B \leqq 200$ mm Kombinationen: $B \leqq 500$ mm	0,2...50 µm (mit verschiedenen Materialien)
Gitter-Spektral-apparate	$A = N\cdot q$ N = Gesamtstrichzahl B = Gitterbreite n = Furchenzahl/mm $N = B\cdot\tilde{n}$	$\Delta\lambda_D = \lambda/q$ q = Ordnung	$B \lesssim 200$ mm $\tilde{n} \lesssim 2\,400\ \mathrm{mm}^{-1}$ $N \lesssim 3\cdot10^5$ $A \lesssim 2\cdot10^6 \ldots 7\cdot10^5$	$q = 1\ldots3$ $\Delta\nu_D \approx 2\cdot10^{14}\ \mathrm{s}^{-1}$ ($q = 3,\ \lambda = 550$ nm)
Fabry-Perot-Inter-ferometer	$A = \pi\frac{\sqrt{B}}{1-R}q = \tilde{N}q$ N = Zahl der miteinander interferierenden Strahlen (Finesse) R = Reflexionsfaktor	$\Delta\lambda_D = \lambda/q$ $\Delta\nu_D = c/2L$ q = Ordnung ν = Frequenz L = Spiegelabstand	$R \approx 0{,}98,\quad \tilde{N} \approx 100$ $L \approx 0{,}3\ldots500$ mm $A \lesssim 10^6\ldots10^8$	$q \approx 5\cdot10^4\ldots10^6$ $\Delta\nu_D \approx 10^{10}\ldots5\cdot10^8\ \mathrm{s}^{-1}$
Spektral-apparate mit durch-stimm-barem Laser	abhängig von Linienbreite des Lasers	abhängig vom Durch-stimmbereich	$\Delta\nu \approx 10^3\ldots10^7$ Hz $A \approx 10^6\ldots10^{11}$	

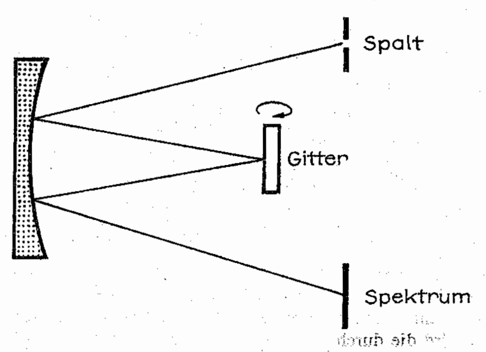


Abb. 7.5.2–15 Plangitterspektograf mit Gitteraufstellung nach EBERT (Seitenansicht)

Quarzträger aufgedampft sind. Es ist daher üblich, billigere Kopien (Replica) der Gitter durch eine Abdrucktechnik herzustellen. Spezielle Werte sind in Tab. 7.5.2–14 angegeben.

Holografische Gitter. Gitter mit großer Lichtstärke und hohem Auflösungsvermögen werden holografisch hergestellt. Dazu wird eine polierte ebene Glasplatte (Ebenheit $< \lambda/4$), die mit einer dünnen fotoempfindlichen Schicht konstanter Dicke bedeckt ist, der Strahlung von zwei kohärenten ebenen Wellen eines Lasers ausgesetzt. Es entsteht in der Fotoschicht ein Interferenzfeld, das durch den Entwicklungsprozeß fixiert wird. Nach Herauslösen des nichtbelichteten Fotoresists entstehen periodische Furchen, deren Kurvenprofil durch eine Sinusfunktion beschrieben wird. Holografisch erzeugte Gitter sind mit Linienzahlen von 2000 bis 4000 Linien/mm bei Abmessungen von 150 mm × 150 mm herstellbar. Sie haben eine hohe Blaze-Effektivität von > 50 % in der 1. Ordnung und sind weitgehend aberrationsfrei herstellbar. Außerdem weisen sie ein geringeres Streulicht als konventionell geteilte Gitter auf.

Typen von Gitterspektralapparaten. *Plangitter-Spektrografen.* Gitterspektrografen werden meist mit Plangittern ausgestattet. Eine häufig verwendete Gitteranordnung geht auf EBERT zurück (Abb. 7.5.2–15): Das vom Eintrittsspalt ausgehende Licht wird vom unteren Teil (Kollimatorbereich) eines großen sphärischen Hohlspiegels parallel gerichtet und auf das Beugungsgitter reflektiert. Die vom Gitter abgebeugte Strahlung gelangt auf den oberen Teil des gleichen Spiegels. Sie wird in der Brennebene des Spiegels (Plattenebene) vereinigt, so daß dort das Beugungsspektrum entsteht. Spektrum und Eintrittsspalt liegen in der gleichen Ebene übereinander. In der Gitteraufstellung nach CZERNY-TURNER werden zwei getrennte sphärische Spiegel verwendet, um einen axialen Strahlengang zu erreichen (Abb. 7.5.2–16).

Konkavgitter-Spektrografen. Konkavgitter bewirken neben der spektralen Zerlegung eine optische Abbildung. Es kann auf zusätzliche abbildende Elemente verzichtet werden. Das ist besonders günstig in Spektralbereichen (Vakuumultraviolett), wo es schwierig ist, Spiegel mit hohen Reflexionsfaktoren herzustellen. Eintrittsspalt, Konkavgitter und Fotoplatte werden dabei auf einem Kreis montiert, der das Gitter berührt (Abb. 7.5.2–17). Der Durchmesser dieses *Rowland-Kreises* muß gleich dem Krümmungsradius des Gitters sein (Typischer Wert: 1 bis 7 m). Bei der *Runge-Paschen-Gitteraufstellung* sind Eintrittsspalt und Gitter fest montiert, während die Fotoplatte entsprechend dem gewünschten Spektralbereich auf dem Rowland-Kreis verschoben werden kann.

Interferenzspektralapparate. Wegen ihrer hohen Auflösung und großen Lichtstärke werden vorzugsweise Vielstrahlinterferenzspektrometer angewendet, deren wichtigster Typ das *Fabry-Perot-Interferometer* (vgl. 7.5.2.1.) ist. Die früher viel verwendeten Lummer-Gehrke-Platten und Stufengitter haben wegen ihrer aufwendigen Herstellung nur noch historisches Interesse. Sie werden kommerziell nicht mehr hergestellt. Fabry-Perot-Interferometer werden zur absoluten Frequenzmessung, als schmalbandige Monochromatoren und hochauflösende Spektrometer (z. B. für das Ausmessen des Linienprofils von Spektrallinien) eingesetzt. Das Auflösungsvermögen beträgt $A = \lambda/\Delta\lambda = q\tilde{N}$, wobei q die Ordnung der Interferenzen (typisch $q \approx 10^4$ bis 10^6) und $\tilde{N}$ die effektive Zahl der miteinander interferierenden Bündel (Finesse $\tilde{N} \approx 50$ bis 100) bedeuten. Das Dispersionsgebiet $\Delta\nu \approx c/2l$ (für $q \gg 1$, $\cos\Theta \approx 1$) nimmt mit wachsendem Spiegelabstand l ab; das Auflösungsvermögen wegen $q \approx 2l/\lambda$ zu. Bei der Untersuchung eines breiten Frequenzspektrums überlagern sich daher die Interferenzringe, die von unterschiedlichen monochromatischen Wellen herrühren. Für ihre Trennung muß ein zusätzlicher Prismenspek-

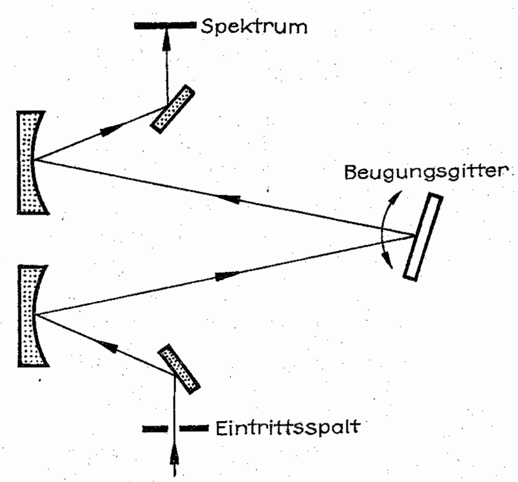


Abb. 7.5.2–16 Plangitterspektograf mit Gitteraufstellung nach CZERNY-TURNER (Aufsicht)

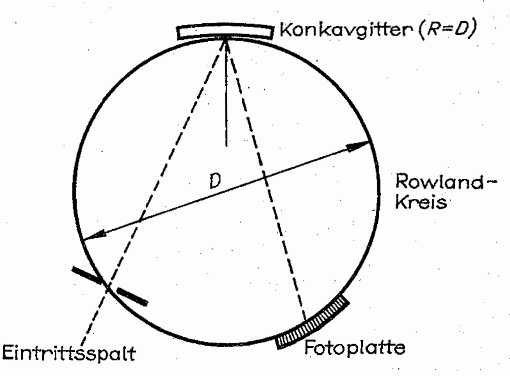


Abb. 7.5.2–17 Runge-Paschen-Gitteraufstellung auf Rowland-Kreis

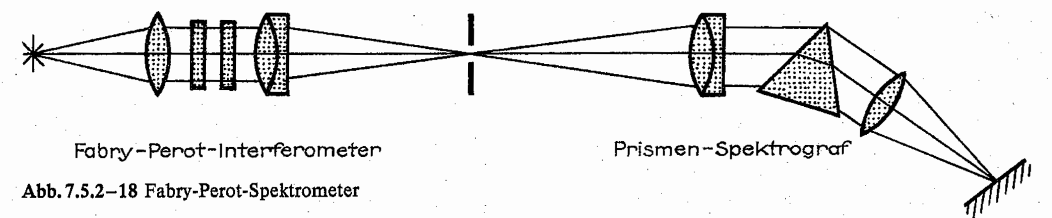


Abb. 7.5.2–18 Fabry-Perot-Spektrometer

trograf nachgeschaltet werden (Abb. 7.5.2–18). In der Registrierebene des Prismenspektrografen treten so viele Spaltbilder auf, wie verschiedene Spektrallinien in dem Frequenzgemisch enthalten sind. In dem vom Eintrittsspalt ausgeschnittenen Ringsystem des Fabry-Perot-Interferometers können feine spektrale Einzelheiten jeder Spektrallinie (Feinstruktur) aufgelöst werden.

Bei gleichem Auflösungsvermögen gelangt beim Fabry-Perot-Interferometer eine höhere Strahlungsleistung auf den Empfänger als bei Gitterspektrometern. Außerdem treten keine „Gittergeister" auf.

Bei einem anderen Anwendungsfall des Fabry-Perot-Interferometers, der insbesondere für den Einsatz als schmalbandiges lichtstarkes Frequenzfilter (Monochromator) geeignet ist, arbeitet man mit einem achsenparallelen (koaxialen) Parallellichtbündel ($\Theta \approx 0$). Die Frequenz der hindurchgehenden Strahlung läßt sich dann mit der Größe des optischen Weges nl zwischen den Spiegelplatten durchstimmen. l kann piezoelektrisch, n durch Änderung des Gasdruckes geändert werden. Für schmalbandige Monochromatoren werden häufig sphärische Spiegel eingesetzt. Das Auflösungsvermögen läßt sich durch Hintereinanderschalten von Fabry-Perot-Interferometern erhöhen.

Fourier-Spektrometer sind Spektralapparate, die auf dem Prinzip der Zweistrahlinterferometrie beruhen. Ihr Wirkprinzip ist folgendes: Wird ein Michelson-Interferometer, dessen Justierung die Beobachtung von Kurven gleicher Neigung in der Beobachtungsebene erlaubt, mit einer monochromatischen Strahlungsquelle der Wellenlänge λ und Intensität S_0 beleuchtet, so ändert sich die Intensität im Zentrum des Ringsystems in Abhängigkeit von der Verschiebung Δs eines Spiegels S_1 nach einer Cosinusfunktion (Abb. 7.5.2–19).

$I(\Delta s) = S_0/2[1 + \cos(4\pi/\lambda)\,\Delta s]$. Enthält die Strahlung der Beleuchtungsquelle dagegen ein diskretes oder kontinuierliches Frequenzspektrum, so erfolgt in Abhängigkeit von Δs eine zusätzliche *Intensitätsmodulation* von $I(\Delta s)$. Bei zwei diskreten Frequenzen ν_1 und ν_2 tritt z. B. eine Schwebung auf, deren Maxima um den Betrag $\Delta s \doteq c/(\nu_1 - \nu_2)$ auseinanderliegen. Mathematisch gesehen stellt das von Δs abhängige Interferenzmuster die *Fourier-Transformierte* der spektralen Verteilung des Lichtes dar, welches das Michelson-Interferometer beleuchtet. Daher kann durch eine *Fourier-Analyse* des registrierten Interferogramms mittels eines Computers auch die spektrale Verteilung der beleuchteten Strahlung berechnet werden. Das *Auflösungsvermögen* des Fourier-Spektrometers hängt vom Verhältnis der maximal möglichen Spiegelverschiebung Δs_{max} zur mittleren Wellenlänge $\bar{\lambda}$ des Spektrums ab und nimmt mit Δs_{max} zu: $A = \lambda/\Delta\lambda = \Delta s_{max}/0{,}35\,\bar{\lambda}$.

Da die Strahlung räumlich nicht zerlegt wird, haben diese Spektrometer eine große Lichtstärke. Sie werden vorzugsweise im Infrarot-Bereich eingesetzt.

Lineare Laserspektrometer. Bei Verwendung von frequenzdurchstimmbaren Lasern lassen sich Absorptions- und Fluoreszenzspektrometer mit hoher Empfindlichkeit und großem Auflösungsvermögen aufbauen. Eine zusätzliche spektrale Zerlegung ist bei diesen Spektrometern in der Regel nicht erforderlich. Bei einem *Transmissions-Spektrometer* wird die durch die Probe hindurchgehende Leistung in Abhängigkeit von der Wellenlänge gemessen. Wird nur ein kleiner Teil ($\lesssim 1\,\%$) der Leistung von der Probe absorbiert, so ist es günstiger, die absorbierte Energie direkt zu messen. Eine Möglichkeit bietet die *fotoakustische Absorptionsmessung*, bei welcher der mit der Absorption verbundene Temperaturanstieg über die entstehende Druckwelle in einem Mikrofon registriert wird. Eine gleichfalls empfindliche Methode beruht auf dem Nachweis der *Fluoreszenzstrahlung*, die mit bestimmten Absorptionsübergängen verknüpft ist.

Diese Methoden erfordern nur geringe Laserleistungen, so

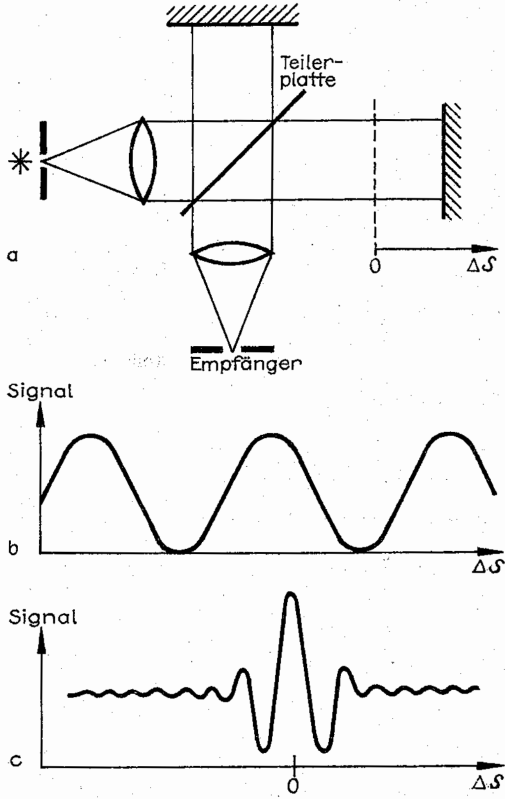


Abb. 7.5.2–19 Fourier-Spektrometer, (a) Aufbau, (b) monochromatisches Signal, (c) Spektrum einer Gasentladungslampe

Tab. 7.5.2–20 Typische wellenlängendurchstimmbare Laser (cw = kontinuierlicher Betrieb, pw = gepulster Betrieb)

Typ	Betriebsart	Arbeitsbereich	Linienbreite	typische Leistung	minimale Impulsbreite
Farbstofflaser	pw	0,28...1,2 μm		0,01...2 MW	≳ 0,1 ps
	cw		< 1 kHz	0,1...1 W	
optisch-parametrisch	pw			1 MW	
als Oszillator		0,4...11 μm	10 MHz...1 GHz		
als Verstärker				≳ 10 MW	≳ 1 ps
Farbzentrenlaser	pw	0,8...1,78 μm	≦ 1 MHz	≈ 10 mW	≳ 10 ns
	cw	2,2...3,3 μm			
Halbleiter-Diodenlaser	pw	0,65...35 μm	< 30 kHz	≈ 10 W	> 3 ps
	cw			1 mW	
Spin-Flip-Laser	pw	5... 6 μm	< 30 MHz	1...10^3 W	≈ 100 ns
	cw	9...14 μm			
Hochdruck-CO_2-Laser	pw	9...11 μm	≈ 300 MHz	> 1 MW	> 30 ps
Differenzfrequenz-Erzeugung	pw	70 μm...1 cm	≈ 300 MHz	1...10^4 W	100 ns

daß nichtlineare Effekte durch Änderung der Besetzungsverteilung der atomaren Energieniveaus vermieden werden (daher: *Lineare* Laserspektroskopie). In Tab. 7.5.2–20 sind typische frequenzdurchstimmbare Laser zusammengestellt. Besondere Bedeutung haben Infrarotspektrometer mit durchstimmbaren Halbleiterlasern erlangt.

Höchstauflösende Laserspektrometer. Mit den klassischen Spektralapparaten und den Verfahren der linearen Laserspektroskopie wird die kleinste noch spektroskopisch nachweisbare atomare Energiedifferenz durch die Linienbreite des atomaren Übergangs begrenzt. Die geringsten Linienbreiten haben Gase unter niedrigem Druck (0,1 bis 10 Pa). Ihre Spektrallinien werden durch den Doppler-Effekt verbreitert, der bei der Absorption bzw. Emission der sich schnell bewegenden Atome auftritt. Typische Doppler-Linienbreiten betragen $\Delta\nu_D \approx (10^7$ bis $10^9)$ Hz. Dieser Wert ist noch groß gegenüber der durch Strahlungsdämpfung bedingten natürlichen Linienbreite von $\Delta\nu_N \approx (10$ bis $10^7)$ Hz. Mit Hilfe eines nichtlinearen Laserspektrometers ist eine Spektroskopie innerhalb der durch den Doppler-Effekt verbreiterten Linie möglich. Ein dafür einsetzbares Spektrometer ist das *Lasersättigungsspektrometer*: Ein Laserbündel großer Intensität (Pumpstrahl) und geeigneter Frequenz wird durch eine mit Gas unter niedrigem Druck gefüllte Küvette hindurchgestrahlt, so daß die Absorptionslinie inhomogen gesättigt wird. Die Breite des in das Doppler-Profil eingebrannten Loches wird durch die natürliche Linienbreite bestimmt. Mit einer zweiten Welle niedriger Intensität (Teststrahl), wird das Loch spektral abgetastet.

Ultrakurzzeit-Spektrometer. Für die Untersuchungen schnell ablaufender Prozesse im Bereich der Mikrophysik, Chemie und Biologie sind Spektralapparate erforderlich, die zeitaufgelöste Messungen im Nanosekunden- (1 ns = 10^{-9} s) und Pikosekundenbereich (1 ps = 10^{-12} s) ermöglichen. Diese Ultrakurzzeit-Spektrometer arbeiten in zwei aufeinanderfolgenden Schritten: Durch einen intensiven kurzen Laserimpuls (*Pumpimpuls*), dessen Impulsbreite klein gegenüber der Meßzeit sein soll, wird der Probe Energie zugeführt, wodurch sie in einen energetischen Nichtgleichgewichtszustand versetzt wird. Die Rückkehr in das thermodynamische Gleichgewicht erfolgt anschließend durch Ausgleichsvorgänge, welche durch die zeitabhängige Registrierung eines charakteristischen Parameters, z. B. der Intensität des Fluoreszenzlichtes, des spektralen Absorptions- und Reflexionsvermögens oder der Brechzahl, gemessen werden können. Während das Fluoreszenzlicht direkt registriert werden kann, erfolgt die Messung der Absorptions-, Reflexions- bzw. Brechzahländerung mit einem zweiten Impuls geringerer Intensität (*Testimpuls*), der gegenüber dem Pumpimpuls zeitlich verzögert eingestrahlt werden kann. Die Zeitverzögerung wird durch eine Verzögerungsstrecke erreicht, in welcher der Testimpuls einen zusätzlichen Weg zurücklegt. Eine Laufstrecke von 0,3 mm ruft eine Zeitverzögerung von 1 ps hervor. Für den Nachweis werden im ns- und Sub-ns-Bereich schnelle fotoelektrische Empfänger eingesetzt. Mit *Schmierkameras* (Streak-Kameras) können noch optische Signale von einigen Pikosekunden Zeitdauer zeitaufgelöst registriert werden. Durch Einsatz *optischer Korrelatoren* (vgl. 7.5.2.3.) lassen sich Signale bequem im Zeitbereich von 1 ns bis einige Femtosekunden (1 fs = 10^{-15} s) vermessen. Als Strahlungsquellen werden im ns- und Sub-ns-Bereich klassische Blitzlampen oder Impulslaser (z. B. Stickstofflaser, gepulste Festkörperlaser mit Güteschaltung) eingesetzt. Für die Erzeugung von Impulsen < 1 ns Impulsdauer werden spezielle Laser im Regime der Modensynchronisation betrieben. Mit ihnen lassen sich Einzelimpulse bzw. Impulsfolgen mit Impulsbreiten von 100 ps bis 100 fs generieren.

7.5.2.3. Optische Korrelatoren. In den letzten Jahren sind die Methoden der Korrelationsmeßtechnik, die sich in der modernen Elektronik gut bewährt haben, auch in der optischen Meßtechnik verstärkt eingesetzt worden. Unter der Korrelation von zwei (optischen) Signalen versteht man den Mittelwert des Produktes dieser Signale. Bei einer zeitlichen Darstellung der Signalgrößen $x(t)$ und $y(t)$ bestimmt man die nichtnormierte *Korrelationsfunktion* $K(\tau)$ in Abhängigkeit von der zeitlichen Verzögerung τ eines Signals durch $K(\tau) = \overline{x(t) \cdot y(t-\tau)}$, wobei der Querstrich die zeitliche Mittelung bedeutet. Wenn $x(t) = y(t)$ ist, spricht man von einer *Autokorrelationsfunktion* (*AKF*) $G(\tau) = \overline{x(t) \cdot x(t-\tau)}$, anderenfalls von der *Kreuzkorrelationsfunktion* (*KKF*) $K(\tau)$. Mit Hilfe der Korrelationsmeßtechnik

können sowohl Signale, deren Funktionswerte eindeutig von einem Argument abhängen, als auch statistische Signale vermessen werden. Der schematische Aufbau eines *Korrelators* ist in Abb. 7.5.2–21 dargestellt. Er besteht aus dem Multiplikator X, dem Verzögerungsglied V und dem Integrator I, der die Mittelwertbildung vornimmt. Je nach Einsatz können Multiplikation und Verzögerung sowohl optisch als auch nach Umwandlung der optischen Signale in elektrische Signale elektronisch erfolgen. Optische Korrelatoren verwendet man beispielsweise in Verfahren zur Impulsbreitenmessung im ps-Gebiet, zur Messung der spektralen Leistungsdichte und zur Bestimmung der Kohärenzeigenschaften höherer Ordnung der optischen Strahlung. Den einfachsten optischen Korrelator stellt das *Fourier-Spektrometer* dar, bei dem zunächst die *Feldstärke-AKF* bestimmt wird, aus der sich mittels Fourier-Transformation die spektrale Leistungsdichte berechnen läßt. Die *KKF* wird bei den *Sampling-Verfahren* gebildet. Sie gestatten es, Impulsfolgen hoher Folgefrequenz mit Empfängern auszumessen, die große Zeitkonstanten aufweisen. Die Ermittlung der *AKF* von statistischen Größen erfolgt in dem Intensitätsinterferometer von Hanbury-Brown und Twiss (vgl. 7.5.3.7.).

Abb. 7.5.2–21 Optischer Korrelator

7.5.3. Messung physikalischer Größen optischer Strahlung

Die elektromagnetische Strahlung im Wellenlängenbereich von $\lambda > 10$ nm (Röntgen-Gebiet) bis 1 mm (Mikrowellengebiet) wird optische Strahlung genannt. In diesem Spektralgebiet müssen die Wellen- und Quanteneigenschaften der elektromagnetischen Strahlung gleichwertig berücksichtigt werden. Die Behandlung der optischen Strahlung als *elektromagnetische Welle* ist besonders geeignet, wenn Ausbreitungserscheinungen untersucht werden. Ihre Beschreibung durch einen Strom von Lichtquanten, den *Photonen*, wird bei der Wechselwirkung der Strahlung mit Materie (z. B. fotoelektrischer Effekt, Wirkungsweise der Laser) und bei der Untersuchung von Schwankungserscheinungen angewendet. Die Quanteneffekte nehmen mit steigender Frequenz zu. Wichtige Parameter zur Beschreibung der optischen Strahlung sind Wellenlänge, Frequenz, Lichtgeschwindigkeit, Phase, Polarisationszustand, Strahlungsleistung und Kohärenz.

7.5.3.1. Wellenlänge. Die in der elektromagnetischen Strahlung enthaltenen Wellenlängen bzw. Spektralanteile werden mit Spektrometern gemessen. Je nach erforderlicher Meßgenauigkeit und auszumessendem Spektralgebiet werden Prismen-Spektrometer, Gitterspektrometer oder Interferometer eingesetzt (vgl. 7.5.2.2.). Die genauesten Messungen erfolgen interferometrisch. Mit speziellen Fabry-Perot-Interferometern sind Wellenlängenbestimmungen mit Unsicherheiten bis $\approx \pm 10^{-10}$ erreichbar. Dazu ist es erforderlich, die Wellenlänge auf das Vakuum zu beziehen. Neben der Temperatur des Interferometers müssen die Größen bekannt sein, welche die Brechzahl der Luft und damit die Lichtgeschwindigkeit beeinflussen: Luftdruck, Lufttemperatur und Luftfeuchte. Eine derartige hohe Meßgenauigkeit ist aber nur sinnvoll, wenn die Strahlungsquelle genügend schmalbandig und frequenzstabil ist. Das gilt annähernd für ausgewählte atomare Übergänge in besonders konstruierten Gasentladungslampen. Mit frequenzstabilisierten Gaslasern sind jedoch wesentlich höhere Frequenzstabilitäten erreichbar, die aber für absolute Wellenlängenmessungen wegen der Unsicherheit in der bisherigen Längendefinition nicht ausgenutzt werden konnten. Bis zum Oktober 1984 beruhte die Definition der Längeneinheit „Meter" auf einer Wellenlängenmessung eines bestimmten atomaren Fluoreszenzüberganges. Danach war das Meter das 1650763,73fache der vom Krypton-Isotop Kr-86 für einen bestimmten Übergang im Vakuum emittierten Wellenlänge. Bei der Vermessung mit frequenzstabilisierten Lasern zeigte es sich, daß diese rosarote Linie nur mit einer mittleren Abweichung von $\pm 4 \cdot 10^{-9}$ darstellbar ist. Bei speziellen frequenzstabilisierten Lasern werden jedoch Frequenzstabilitäten mit Unsicherheiten $\Delta\nu/\nu \lesssim 10^{-14}$ erreicht. Diese hohe Frequenzstabilität ließ sich aber wegen der größeren Unsicherheit in der Definition des Längenstandards für die Längenmessungen nicht voll ausnutzen. Eine Abhilfe konnte nur eine Neudefinition der Längeneinheit bringen.

7.5.3.2. Frequenz. Frequenz- bzw. Zeitmessungen werden seit 1967 auf die im Mikrowellengebiet bei $\approx 9{,}2$ GHz liegende Normalfrequenz bezogen. Sie wird in einem Normalfrequenzgenerator (Atomuhr) erzeugt, dessen Schwingungsfrequenz durch einen Übergang zwischen zwei Hyperfeinstrukturniveaus eines Caesium-Isotops Cs-133 festgelegt ist. Die Sekunde wird als das 9192631770fache der Periodendauer dieser Übergangsfrequenz definiert. Caesium-Normaluhren haben Langzeitstabilitäten von 10^{-14} und Reproduzierbarkeiten von 10^{-13}. Ihre Genauigkeit ist daher erheblich größer als die des bisherigen Längenstandards. Direkte Frequenzmessungen mit Anschluß an das Frequenznormal sind jedoch elektronisch nur bis zu Frequenzen von ≈ 1 THz ($\triangleq \lambda = 300$ µm) möglich. Dazu wird die Frequenz eines Klystrons ($\nu \approx 50$ bis 70 GHz) an einer speziellen Metall-Oxid-Metall-Punktkontakt-Diode (MOM-Diode) vervielfacht und mit der unbekannten Frequenz gemischt. Die Differenzfrequenz wird dann direkt mit elektronischen Mitteln (z. B. durch Frequenzzähler) gemessen. Da es nicht möglich ist, Vielfache der Mikrowellenfrequenzen bis in den Infrarot- und sichtbaren Bereich (> 100 THz) zu erzeugen, wird eine Kette von stabilisierten Lasern verwendet, deren Frequenzen durch Erzeugung von Harmonischen bzw. Summen- und Differenzfrequenzen in der MOM-Punktkontaktdiode vermessen werden, und die dann als sekundäre Frequenzstandards zur Verfügung stehen. In dieser Frequenzkette wurden eingesetzt: HCN-Laser (0,80 und 0,89 THz), H_2O-Laser (3,82 und 10,72 THz), CO_2-Laser (≈ 28 THz $\triangleq 10{,}6$ µm) und He-Ne-Laser mit Methan stabilisiert (88,38 THz $\triangleq 3{,}39$ µm). Die höchste bisher

vermessene Frequenz ist die der roten Strahlung des mit Iodgas von I-127 frequenzstabilisierten He-Ne-Lasers (473,61 THz ≙ 633 nm), die mit einer Genauigkeit $\pm 2 \cdot 10^{-10}$ ermittelt wurde.

7.5.3.3. Lichtgeschwindigkeit. Die Vakuumlichtgeschwindigkeit wurde seit ihrer ersten Messung 1675 durch ROEMER immer wieder mit den unterschiedlichsten Methoden und mit erhöhter Genauigkeit gemessen. Die genauesten Verfahren beruhen auf der gleichzeitigen Messung von Frequenz ν und Wellenlänge λ. Die Lichtgeschwindigkeit berechnet sich daraus $c = \nu\lambda$. Derartige Messungen wurden von verschiedenen Laboratorien mit der Strahlung von He-Ne-Lasern bei 3,39 µm durchgeführt, die an einer Absorptionslinie von Methan frequenzstabilisiert waren. Als bester Wert wurde für die Lichtgeschwindigkeit im Vakuum $c = 299\,792\,458{,}0 \pm 1{,}2\ \mathrm{ms^{-1}}$ erzielt. Die Unsicherheit von $\pm 4 \cdot 10^{-9}$ ergab sich allein aus der bisherigen Meter-Definition. Um prinzipiell auch die Länge mit höherer Genauigkeit messen zu können, wurde festgelegt, sie durch eine Zeitmessung an das Frequenznormal anzubinden. Seit dem 20. Oktober 1984 wird die Längeneinheit wie folgt definiert: „Das Meter ist die Länge der Strecke, die Licht im Vakuum während der Zeit von 1/299792458 Sekunden zurücklegt". Die Lichtgeschwindigkeit erhält dadurch als universelle Naturkonstante per Definition einen konstanten Wert $c = 299\,792\,458\ \mathrm{ms^{-1}}$. Sie braucht nicht länger experimentell ermittelt zu werden.

7.5.3.4. Phase. Die Phasenänderung einer Lichtwelle nach Durchlaufen eines bestimmten optischen Weges wird interferometrisch gemessen, indem sie mit einer kohärenten Teilwelle (Referenzwelle) überlagert wird, die durch Teilung der Amplitude oder der Wellenfront aus der Meßwelle gewonnen wird. Die möglichen Meßverfahren sind im Abschnitt 7.5.2.1. behandelt.

7.5.3.5. Polarisationszustand. Elektromagnetische Strahlung kann linear, zirkular oder elliptisch polarisiert sein. Mit kohärenten Strahlungsquellen (Laser) läßt sich direkt optische Strahlung in einem bestimmten Polarisationszustand erzeugen. Bei thermischen Lichtquellen werden dagegen Wellenzüge emittiert, deren Polarisationszustand räumlich und zeitlich stark schwankt, so daß ein Strahlungsempfänger im Mittel keinen bevorzugten Polarisationszustand registriert, und die Strahlung unpolarisiert erscheint. Die Messung des Polarisationszustandes einer elektromagnetischen Welle kann mittels eines Polarisators (Analysators) und einer $\lambda/4$-Phasenplatte erfolgen. Man läßt die Strahlung auf den drehbaren Analysator fallen. Folgende drei Fälle sind möglich:

a) Tritt bei einer Stellung des Analysators vollkommene Auslöschung auf, so ist die Welle *linear* polarisiert.

b) Wechselt die Intensität bei der Drehung, ohne daß die Intensität ganz auf Null zurückgeht, so ist die Welle *teilweise linear* polarisiert oder *elliptisch* polarisiert oder *teilweise elliptisch* polarisiert.
In der Minimumstellung des Analysators steht seine Haupttransmissionsrichtung senkrecht zur Schwingungsrichtung der teilweise linear polarisierten Welle oder senkrecht auf der großen Halbachse der Schwingungsellipse.
Es wird dann zusätzlich eine $\lambda/4$-Phasenplatte (oder ein entsprechender Kompensator) senkrecht zur Wellennormalen in den Strahlengang gebracht, wobei deren eine Hauptschwingungsrichtung senkrecht zur Haupttransmissionsrichtung des Analysators (in bisheriger Minimumstellung) liegen soll. Sie fällt dann mit der Schwingungsrichtung des linear polarisierten Teiles bzw. mit der großen Hauptachse der Ellipse zusammen.
Liegt das Intensitätsminimum nach Einsetzen der $\lambda/4$-Phasenplatte bei der gleichen Analysatorstellung, so ist die Welle *teilweise linear* polarisiert.
Ist es erforderlich, den Analysator nach Einsetzen der $\lambda/4$-Phasenplatte um einen Winkel $\pm\psi$ zu drehen, damit die Intensität wieder zu einem Minimum wird, so ist die Welle ganz oder teilweise elliptisch polarisiert. Durch die $\lambda/4$-Phasenplatte ist die Welle in eine ganz oder teilweise linear polarisierte Welle umgewandelt worden, deren Schwingungsrichtung den Winkel $\pi/2 \pm \psi$ gegen die Haupttransmissionsrichtung des Analysators bildet. Der Drehwinkel ψ gibt das Hauptachsenverhältnis $b/a = \tan\psi$ der Schwingungsellipse an. Bei

Tab. 7.5.3-1 Strahlungsempfänger für optische Strahlung

Spektralbereich	Wellenlänge	Strahlungsempfänger	Wirkung der optischen Strahlung
Vakuum-Ultraviolett	10...200 nm	Zählrohr	Ionisation von Gasen
		Szintillationszähler	Lumineszenz und SEV
	ab 150 nm	Schumann-Platte	fotochemischer Prozeß
Ultraviolett	200...400 nm		
		Fotoplatte	fotochemischer Prozeß
sichtbares Gebiet	400...760 nm	Hochvakuum-Fotozelle	
			äußerer Fotoeffekt
		SEV	
nahes Infrarot	0,76...2,5 µm	Halbleiterempfänger	innerer Fotoeffekt
	bis 1,2 µm	Bolometer	
	bis 1,4 µm	Thermoelement	thermische Wirkung
		pyroelektrischer Empfänger	
mittleres Infrarot	2,5...50 µm	Golay-Zelle	
		MOM-Diode	Gleichrichterwirkung
fernes Infrarot	50 µm...1 mm		

völliger Auslöschung der Intensität ist die Welle *elliptisch*, bei teilweiser Auslöschung *teilweise elliptisch* polarisiert.

c) Bleibt die Intensität beim Drehen des Analysators konstant, so ist die Welle *unpolarisiert* oder *zirkular* oder *teilweise zirkular* polarisiert. Zur Unterscheidung wird wieder zusätzlich eine $\lambda/4$-Phasenplatte in den Strahlengang eingeschaltet. Eine besondere Einstellung ihres Azimutwinkels ist nicht notwendig.
Ändert sich beim anschließenden Drehen des Analysators die Intensität nicht, so ist die Welle *unpolarisiert*. Tritt völlige Auslöschung bei einer Analysatorstellung auf, liegt eine *zirkular-polarisierte* Welle vor; bei partieller eine *teilweise zirkular* polarisierte Welle.

7.5.3.6. Strahlungsleistung. Mit jeder optischen Strahlung ist ein Energietransport verbunden. Die Strahlungsenergie Q_e ist gleich dem Produkt aus Strahlungsleistung Φ_e und Zeit t, $Q_e = \Phi_e t$. Wenn Φ_e zeitabhängig ist, so wird über die Zeit integriert, $Q_e = \int_0^{t_M} \Phi e(t)\,dt$, wobei t_M die Meßzeit darstellt. Absolute Strahlungsenergien und -leistungen werden kalorimetrisch gemessen. Dazu läßt man die Strahlung in einem bestimmten Volumen möglichst vollständig absorbieren und mißt den Temperaturanstieg ΔT. Die absorbierte Strahlungsenergie Q_e beträgt dann $Q_e = c_w m \Delta T$, wobei c_w die spezifische Wärme und m die Masse des Meßkörpers bedeuten. Der absorbierende Meßkörper wird meist so ausgeleuchtet, daß die Strahlung entweder durch Vielfachreflexionen an der Oberfläche gut absorbiert wird oder tief in den Meßkörper eindringen kann, um dort im Volumen in Wärme umgewandelt zu werden. *Strahlungskalorimeter* für Oberflächenabsorption haben meist die Form eines offenen Kegels aus Graphit oder Kupfer (für sehr hohe Strahlungsleistungen), in dessen freie Öffnung eingestrahlt wird. Der Temperaturanstieg wird elektrisch mit Thermistoren in einer Brückenschaltung gemessen, wobei die Thermistoren auf einem nichtbestrahlten Kegel als Vergleichswiderstände dienen. Kegelkalorimeter werden für Strahlungsenergien $\gtrsim 1$ mJ bzw. für Strahlungsleistungen > 1 mW bis ≈ 1 kW gebaut. Niedrigere Strahlungsenergien (bis 1 µJ) lassen sich mit Glasabsorbern bestimmen, deren Oberflächentemperatur gemessen wird. Für den *Strahlungsnachweis* sowie für *relative Leistungs-* und *Energiemessungen* stehen im Spektralbereich der optischen Strahlung eine Vielzahl unterschiedlicher Empfänger zur Verfügung. Wichtige Kenngrößen dieser Empfänger sind der spektrale Empfindlichkeitsbereich, ihre Nachweisgrenze (bedingt durch das Signal-Rausch-Verhältnis) und ihre Zeitkonstante, welche die kleinste meßbare Zeitdauer von Strahlungsimpulsen bzw. zeitlichen Schwankungen des Strahlungsflusses angibt. Wichtige Strahlungsempfänger sind in Tab. 7.5.3–1 zusammengestellt.

Generell lassen sich Empfänger für optische Strahlung einteilen in *Quantendetektoren* und *thermische Empfänger*. Quantendetektoren sprechen auf die Energie der Photonen an. Ihre Strahlungsempfindlichkeit ist daher stark frequenzabhängig. Sie haben eine große Empfindlichkeit und kleine Zeitkonstanten. Bei den thermischen Empfängern wird wie bei den Kalorimetern die Strahlung in einer Absorberschicht in Wärme umgewandelt und anschließend die Temperatursteigerung gemessen. Ihre Zeitkonstante ist umso niedriger, je kleiner ihre Wärmekapazität ist. Ihre Empfindlichkeit ist bei geeigneter Absorberwahl weitgehend frequenzunabhängig.

Die *fotografische Schicht* kann als Strahlungsempfänger im Wellenlängenbereich von > 10 nm bis $\approx 1{,}5$ µm eingesetzt werden. Sie enthält meist Silberhalogenid-Emulsionen, die auf Glasplatten oder Filme aufgetragen sind. Für spezielle Spektralbereiche müssen sie besonders sensibilisiert werden. Die durch die Strahlung und nachfolgende Entwicklung erzeugte Schwärzung kann mit Hilfe von Fotometern quantitativ ausgemessen werden. Nach entsprechender Eichung können Fotoplatten zur Messung von Strahlungsenergien eingesetzt werden. Sie sind insbesondere für sehr niedrige Strahlungsleistungen bzw. -energien geeignet.

In den *Hochvakuum-Fotodioden* und *Sekundärelektronenvervielfachern* (SEV) wird die Strahlung von der Fotokatode absorbiert. Sie emittiert aufgrund des äußeren Fotoeffektes Elektronen, sofern die Photonenenergie $h\nu$ (h = Plancksches Wirkungsquantum, ν = Frequenz) größer als die Austrittsarbeit des Materials ist. Die Zahl der ausgelösten Elektronen ist proportional der auffallenden Strahlungsleistung. Der gesamte Spektralbereich, der mit unterschiedlichen Katodenmaterialien empfangen werden kann, erstreckt sich von > 100 nm bis $\sim 1{,}2$ µm. Bei der Fotodiode werden die austretenden Elektronen von der positiv vorgespannten Anode gesammelt und erzeugen einen Fotostrom. Fotodioden haben eine kleine Zeitkonstante (< 1 ns) und einen großen Aussteuerbereich bei relativ geringer Empfindlichkeit. In dem SEV gelangen die Elektronen nacheinander zu einer Reihe positiv vorgespannter Elektroden, den Dynoden. Hier erzeugt jedes auftreffende Elektron ~ 2 bis 3 Sekundärelektronen, die von der folgenden Dynode gesammelt werden. Dadurch wird eine innere rauscharme Verstärkung des Fotostromes um den Faktor 10^5 bis 10^8 erzielt, bevor die Elektronen auf die Anode treffen. Der SEV zeichnet sich daher durch eine sehr große Nachweisempfindlichkeit aus, die durch den Dunkelstrom begrenzt wird. Bei genügend kleinen Strahlungsleistungen können mit einer Fotodiode oder einem SEV direkt Einzelphotonen gezählt werden.

Halbleiterempfänger nutzen den inneren Fotoeffekt aus. Durch Absorption der Photonen wird die Leitfähigkeit der Materialien erhöht, indem Ladungsträger entweder aus dem Valenzband in das Leitungsband angeregt oder aus Störstellen freigesetzt werden. Die langwellige Grenzwellenlänge wird durch den Bandabstand oder den energetischen Abstand der Störstellen von einem Band bestimmt. Diese Fotoleiter sind vom sichtbaren Spektralbereich bis ins FIR einsetzbar. Von besonderer Bedeutung ist ihr Einsatz für > 1 µm, da dort der Quantenwirkungsgrad der Fotokatoden gegen Null geht. Mit Fotodioden werden Zeitkonstanten < 1 ns erreicht. Fotoleiter im IR-Gebiet erfordern meist eine Kühlung mit flüssigem Stickstoff oder Helium.

In *thermischen Empfängern* wird die Strahlungsenergie in der Absorberfläche in Wärme umgewandelt. Die dadurch hervorgerufene Temperatursteigerung wird durch die Änderung des elektrischen Widerstandes *(Bolometer)*, die Erzeugung einer Thermospannung *(Thermoelement)* oder Ausdehnung eines Gasvolumens *(Golay-Zelle)* gemessen. *Supraleitende Bolometer* nutzen die starke Zunahme des elektrischen Widerstandes beim Übergang eines Supraleiters in den normalleitenden Zustand aus. Dadurch läßt sich die Empfindlichkeit um mehr als zwei Größenordnungen steigern. *Vakuumthermoelemente* werden insbesondere zum Nachweis von IR-Strahlung eingesetzt. Bei Erwärmung einer Lötstelle des Thermoelementes durch die auffallende Strahlung tritt eine Thermospannung auf. Sie kann durch Hintereinanderschalten mehrerer Elemente erhöht werden

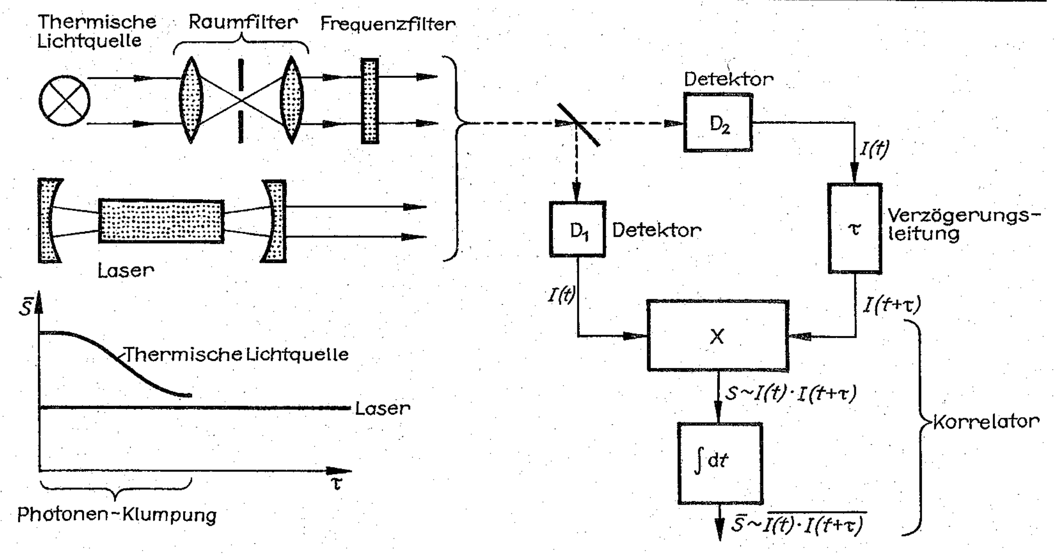


Abb. 7.5.3–2 Intensitätsinterferometer

(Thermosäule). Die *Golayzelle* ist ein spezieller IR- und FIR-Detektor. Die Strahlung erhitzt in einer Gaskammer eine geschwärzte Membran, die das Kammergas erwärmt. Es kommt zu einer Druckerhöhung im Gas, so daß die Membran sich durchwölbt. Die Durchbiegung kann sehr empfindlich lichtoptisch oder kapazitiv gemessen werden.

7.5.3.7. Kohärenz der optischen Strahlung. Laser im transversalen Grundmodus (TEM_{00}) und thermische Lichtquellen, die beugungsbegrenzt sind, haben bei gleicher spektraler Bandbreite die gleiche *Kohärenz erster Ordnung* (vgl. 7.5.2.1.), d.h. sie unterscheiden sich bei den klassischen Interferenzexperimenten nur durch die wesentlich höhere Ausgangsleistung des Lasers. Prinzipielle Unterschiede zwischen Lasern und thermischen Lichtquellen ergeben sich allerdings bei Experimenten mit dem optischen Korrelator von Hanbury-Brown und Twiss (Abb. 7.5.3.–2). Hierbei wird das Licht der Strahlungsquelle nach der Teilung auf die beiden lichtelektrischen Detektoren D_1 und D_2 gegeben, wobei das elektrische Ausgangssignal eines der Detektoren zeitlich um den Wert τ verzögert werden kann. Die Fotoströme der beiden Detektoren $I(t)$ und $I(t+\tau)$, die proportional der auffallenden Lichtleistung sind, werden in einem Multiplikator elektronisch multipliziert und anschließend zeitlich gemittelt. Das resultierende Signal $G(\tau) = \overline{I(t) \cdot I(t+\tau)}$ stellt die Intensitätsautokorrelationsfunktion des Eingangssignals dar. Sie hängt nun unterschiedlich von τ ab, je nachdem ob es sich um eine thermische Lichtquelle oder um einen Laser handelt. Während das Ausgangssignal bei einer thermischen Lichtquelle sehr stark mit τ abfällt, ist es bei einem Laser nahezu unabhängig von τ. Diese unterschiedliche Verhaltensweise führt zu einem allgemeineren Kohärenzbegriff. Danach ist die Strahlung eines Lasers von *höherer Ordnung kohärent* als die einer thermischen Lichtquelle. Physikalisch heißt das, daß im thermischen Licht die Photonen im zeitlichen Mittel nicht mit konstanten zeitlichen Abständen, sondern in Form von Klumpen auf den Empfänger treffen. Dieser Klumpungseffekt wird auch *Photonbunching* genannt.

7.5.4. Messung optischer Stoffeigenschaften

7.5.4.1. Brechzahl und Dispersion transparenter Medien. In der technischen Optik ist es üblich, die Brechzahl n eines Mediums als Verhältnis der Lichtgeschwindigkeit in

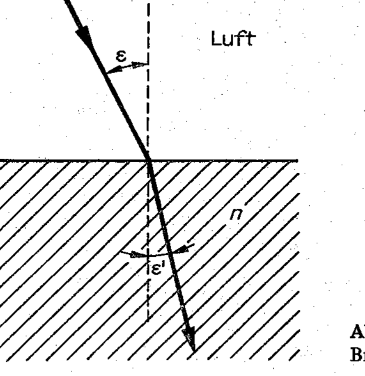


Abb. 7.5.4–1 Brechungsgesetz

Tab. 7.5.4–2 Spektrallinien, die als Wellenlängennormale verwendet werden

Index	Wellenlänge in Luft in nm	angeregtes Element	Spektralfarbe
h	404,66	Hg	Violett
F'	479,99	Cd	Blau
F	486,13	H	Blau
e	546,07	Hg	Grün
d	587,56	He	Gelb
C'	643,85	Cd	Rot
C	656,28	H	Rot

Luft c zur Lichtgeschwindigkeit in dem Medium v_M anzugeben: $n = c/v_M$. Dabei bezieht man sich auf Luft bei 20 °C, einem Druck von 0,1013 MPa und einer Luftfeuchte von ≈ 10 %. Die auf das Vakuum bezogene Brechzahl ist um den Faktor 1,0003 größer. Beim Übergang eines Lichtstrahls aus Luft in ein Medium mit der Brechzahl n (Abb. 7.5.4–1) besteht zwischen dem Einfallswinkel ε und Brechungswinkel ε' der Zusammenhang $\sin\varepsilon' = (1/n)\sin\varepsilon$ *(Brechungsgesetz)*. Beim Übergang von Licht von einem Medium 1 (z.B. Wasser) in ein Medium 2 (z.B. Glas) gilt generell $n_1 \sin\varepsilon_1 = n_2 \sin\varepsilon_2$, wobei n_1, ε_1 sich auf das Medium 1 und n_2, ε_2 auf das Medium 2 beziehen. Da die Brechzahl von der Wellenlänge des Lichtes abhängt, ist bei Angabe von n erforderlich, die zugehörige Wellenlänge zu vermerken. Das geschieht durch einen Index, der sich jeweils auf eine spezielle Spektrallinie bezieht, die als Wellenlängennormal verwendet worden ist (Tab. 7.5.4–2). Zur Charakterisierung optischer Gläser wird die *Hauptbrechzahl* n_e angegeben, die mit dem Licht der grünen Hg-Linie bei $\lambda_e = 546{,}07$ nm ermittelt worden ist. Bei optischen Gläsern liegt die Hauptbrechzahl n_e zwischen 1,3 und 2,1. Die Differenz der Brechzahlen für zwei verschiedene Lichtwellenlängen (d. h. Lichtfarben) wird in der technischen Optik *Dispersion* genannt. Die *Hauptdispersion* $n_{F'}$-$n_{C'}$ gibt die Differenz der Brechzahlen zwischen dem blauen und dem roten Bereich des sichtbaren Spektrums an. Da die Dispersion im blauen Bereich des Spektrums größer ist als im roten, gibt man zusätzlich die *Teildispersionen* an: n_h-n_F, $n_{F'}$-n_F, usw. Die *relative Teildispersion* ϑ ist das Verhältnis von Teildispersion und Hauptdispersion. Sie wird zur Korrektion der Farbfehler höherer Ordnung in Linsensystemen benötigt. Zur Korrektur einfacher Farbfehler genügt die Angabe der *Abbeschen Zahl* $\nu_e = (n_e - 1)/(n_{F'} - n_{C'})$. Für optische Gläser liegt sie im Bereich von 10 bis 120.

Zur Bestimmung der Brechzahlen fester, flüssiger und gasförmiger Proben werden die folgenden Meßverfahren angewendet:

Messung der Strahlablenkung in einem Prisma, das aus der zu messenden Substanz gefertigt wurde. Das Verfahren ist sehr genau aber aufwendig.

Messung des Grenzwinkels der Strahlablenkung. Diese Verfahren sind insbesondere für Routineuntersuchungen geeignet. Ihre Genauigkeit ist nur wenig geringer.

Interferometrische Verfahren haben die höchste Meßgenauigkeit, sind jedoch nicht universell anwendbar.

Verfahren der Strahlablenkung mittels Prismen. Aus der zu messenden Substanz wird ein Prisma gefertigt, das neben der geschliffenen Grundfläche nur zwei exakt plangeschliffene und polierte Begrenzungsflächen aufweisen muß. Der brechende Winkel χ soll optimal ≈ 60° betragen (Abb. 7.5.4–3). Die Vermessung dieses Winkels sowie der Strahlablenkung erfolgt mit einem *Goniometer*, das einem Prismen-Spektrometer entspricht. Es besteht aus einem *Kollimator*, der das Prisma mit einem Parallelstrahlbündel beleuchtet, und dem schwenkbaren *Fernrohr* mit Fadenkreuz. Der *Prismentisch* ist drehbar. Mit dem Hauptteilkreis kann die Drehung des Fernrohres bei Präzisionsgeräten bis auf ± (1″ bis 2″) genau gemessen werden. Die Bestimmung des brechenden Winkels χ des Prismas erfolgt durch Autokollimation. Dazu wird mittels eines *Gaußschen Okulars* das Fernrohr jeweils senkrecht auf die brechenden Flächen eingestellt. Die höchste Meßgenauigkeit zur Bestimmung von n wird durch das Verfahren der *Minimalablenkung* erzielt. Bei symmetrischem Durchgang der Lichtstrahlen durch das Prisma ist die Ablenkung am geringsten. Der Winkel δ der

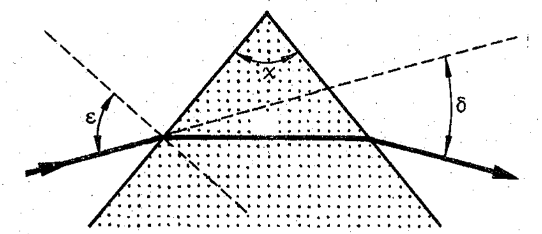


Abb. 7.5.4–3 Brechung am Prisma

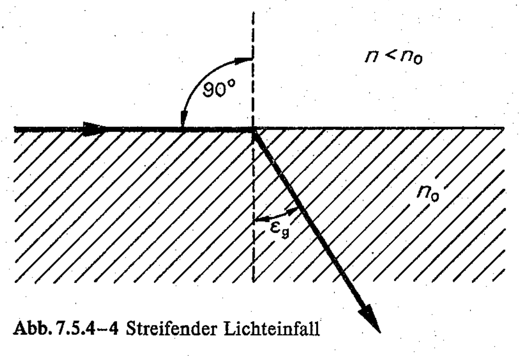


Abb. 7.5.4–4 Streifender Lichteinfall

Minimalablenkung wird bestimmt, indem man beim gleichmäßigen Drehen des Prismentisches die Umkehr der Strahlablenkung beobachtet. Die Brechzahl n errechnet sich dann aus $n = \sin 1/2(\chi + \delta)/\sin 1/2\chi$. Bei einem Fehler von ≈ ±2″ bei der Bestimmung der Prismenwinkel χ, δ kann man eine relative Genauigkeit von $\Delta n/n \approx \pm 1 \cdot 10^{-6}$ erzielen. Die Brechzahlen von Flüssigkeiten oder Gasen können nach dieser Methode ermittelt werden, indem man sie in Hohlprismen einfüllt.

Verfahren der Grenzwinkelbestimmung (Refraktometrie). Beim Übergang eines Strahlenbündels von einem Medium mit niedrigerer Brechzahl n in ein Medium mit höherer Brechzahl n_0 ist der Brechungswinkel ε_g am größten für *streifenden Lichteinfall* (Abb. 7.5.4–4). Aus dem Brechungsgesetz folgt $n \sin 90° = n_0 \sin\varepsilon_g = n$. Die gleiche Beziehung gilt für den umgekehrten Strahlengang. ε_g ist dann der Grenzwinkel der Totalreflexion. Aus der Messung des Grenzwinkels kann die Brechzahl eines Mediums bestimmt werden, wenn die des anderen bekannt ist. Die Beobachtung des Grenzwinkels kann im durchgehenden Licht bei streifendem Einfall oder im reflektierten Licht bei Totalreflexion erfolgen. Die Beleuchtung geschieht mit einer ausgedehnten Lichtquelle. Mit einem Meßfernrohr wird die Hell-Dunkel-Grenze beobachtet.

Spezielle Refraktometer. Beim *Halbkugelrefraktometer* (Abbesches Kristallrefraktometer) ist das hochbrechende Medium (Brechzahl n_0) als Halbkugel ausgebildet (Abb. 7.5.4–5). Auf diesen Kontaktkörper wird der Meßkörper aufgesetzt. Der Grenzwinkel ε_g wird direkt mit einem Meßfernrohr gemessen, indem auf die Hell-Dunkel-Grenze eingestellt wird. Der Einfluß der Kugel auf den teleskopischen Fernrohrstrahlengang wird durch eine plankonkave Linse korrigiert. Diese Korrektur entfällt bei den *Refraktometern mit prismatischen Kontaktkörpern.* Am bekanntesten sind das *Pulfrich-Refraktometer* (rechtwinkliges Prisma) und das *Abbe-Refraktometer* (spitzwinkliges Prisma) (Abb. 7.5.4–6). Beide sind sowohl für die Untersuchung von Flüs-

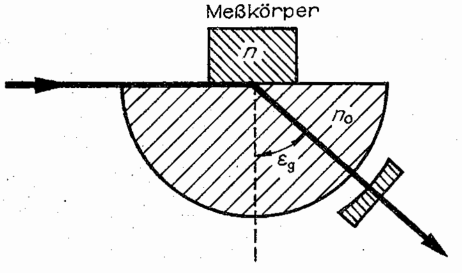


Abb. 7.5.4–5 Halbkugelrefraktometer

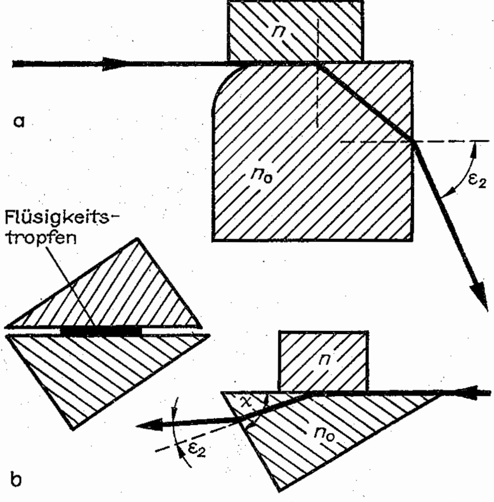


Abb. 7.5.4–6 (a) Kontaktprisma des Pulfrich- und (b) des Abbe-Refraktometers

sigkeiten als auch von festen Körpern geeignet. Feste Prüfkörper werden mit einer höherbrechenden Flüssigkeit aufgeklebt. Beim Abbe-Refraktometer wird dazu das untere Prisma entfernt. Flüssigkeiten werden bei diesem Refraktometer zwischen beiden Prismen vermessen. Die gesuchte Brechzahl n berechnet sich beim Pulfrich-Refraktometer $n = \sqrt{n_0^2 - \sin^2\varepsilon_2}$ und beim Abbe-Refraktometer zu $n = \sin\chi\sqrt{n_0^2 - \sin^2\varepsilon_2} \pm \cos\chi\sin\varepsilon_2$ (± je nachdem, ob der Grenzstrahl am Winkel χ zur brechenden Kante hin oder weggebrochen wird). Der Vorzug dieser Refraktometer besteht in ihrer einfachen Handhabung, relativ großen Genauigkeit und in der geringen erforderlichen Substanzmenge. *Eintauch-Refraktometer* dienen zur Brechzahlbestimmung in ruhenden oder strömenden Flüssigkeiten. Sie sind auch als automatische Anzeigegeräte für die Betriebskontrolle ausgebildet worden.

Interferometrische Verfahren haben die höchste Meßgenauigkeit von allen optischen Meßverfahren. Von Bedeutung für die Brechzahlbestimmung ist das Interferometer von Rayleigh-Löwe-Haber (vgl. 7.5.2.1.).

7.5.4.2. Optische Konstanten absorbierender Medien. Stark absorbierende Medien, z. B. Metalle, werden neben ihrer Brechzahl n durch ihren Absorptionskoeffizienten $n\varkappa$ charakterisiert. Beide optische Konstanten werden formal in einer komplexen Brechzahl $n' = n(1 - i\varkappa)$ zusammengefaßt. Sie werden aus Reflexionsmessungen gewonnen. Damit lassen sich insbesondere Oberflächenschichten sehr genau vermessen. Das wichtigste Meßverfahren beruht auf Winkelmessungen. Bei der Reflexion linear polarisierten Lichtes unter verschiedenen Einfallswinkeln an der zu untersuchenden Oberfläche entsteht elliptisch polarisiertes Licht. Die Parameter der Schwingungsellipse können allein durch Winkelmessungen ermittelt werden (vgl. 7.5.4.3.). Dieses Untersuchungsverfahren wird *Ellipsometrie* genannt. Die Ellipsometrie ist eine der empfindlichsten Methoden zur Bestimmung der optischen Parameter dünner stationärer Oberflächenschichten (Schichtdicke, n, $n\varkappa$) bzw. deren zeitliche Änderung (Schichtwachstum). Unter geeigneten Bedingungen lassen sich Schichtdicken von einigen Atomlagen messen. Derartige dünne Oberflächenschichten spielen eine wichtige Rolle in der Oberflächenphysik und -chemie, insbesondere in der Halbleitertechnologie, in der Werkstoffkunde sowie in der Biologie und Medizin.

Ellipsometer. Das linear polarisierte Licht einer Hg-Hochdrucklampe oder eines He-Ne-Lasers wird in einem Ellipsometer durch die Reflexion an der stark absorbierenden Probe in elliptisch polarisiertes Licht umgewandelt. Die Parameter der Schwingungsellipse des reflektierten Lichtes werden vermessen, indem das elliptisch polarisierte Licht mittels eines Kompensators in linear polarisiertes Licht umgewandelt wird. Die Lage der Schwingungsebene (Azimut) wird mit einem Analysator bestimmt. Dazu werden Polarisationswinkel φ_P und Analysatorwinkel φ_A so eingestellt, daß die Ausgangsintensität Null wird. Aus beiden Winkeln lassen sich die relative Amplitudendämpfung und die relative Phasenverschiebung der parallel und senkrecht zur Einfallsebene schwingenden Feldstärkekomponenten ermitteln, mit denen dann die optischen Konstanten berechnet werden können. Moderne Geräte verfügen über eine fotoelektrische Registrierung. Spitzengeräte erlauben Änderungen der mittleren Filmdicke von 0,007 nm bzw. die Dicke von Monolagen bis zu 0,003 nm zu messen.

7.5.4.3. Doppelbrechung. Zur Bestimmung der Doppelbrechung eines anisotropen transparenten Mediums wird die Phasendifferenz gemessen, die zwischen ordentlichem und außerordentlichem Strahl beim Durchgang durch das Medium auftreten. Der Wert der Doppelbrechung wird durch die größte Differenz der Brechzahlen von außerordentlichem (n_{a0}) und ordentlichem (n_0) Strahl $\Delta n = n_{a0} - n_0$ ausgedrückt. Kristalle mit $\Delta n > 0$ heißen positiv doppelbrechend, mit $\Delta n < 0$ negativ doppelbrechend.

Meßverfahren. Es werden nur doppelbrechende Medien betrachtet, bei denen die Schwächung von ordentlichem und außerordentlichem Strahl gleich groß ist. Die Messung wird mit monochromatischem Licht an einer planparallelen Platte durchgeführt, die so aus dem doppelbrechenden Kristall geschnitten ist, daß die optische Achse in der Eintrittsfläche liegt.

Bestimmung der Lage der Hauptschwingungsrichtungen („schnelle" und „langsame" Achse). Die Kristallplatte wird senkrecht zwischen zwei gekreuzte Polarisatoren gebracht und mit parallelem monochromatischem Licht durchstrahlt. Die Intensität hinter dem Analysator wird in Abhängigkeit vom Drehwinkel der Kristallplatte beobachtet. Die nach Einbringen der Kristallplatte aufgetretene Aufhellung verschwindet wieder, wenn die Achsen der Platte parallel zu den Haupttransmissionsrichtungen der Polarisatoren liegen.

Messung der Phasendifferenz. Die Kristallplatte wird jetzt so gedreht, daß ihr Achsenkreuz gegenüber der Schwingungs-

richtung des linear polarisierten Lichtes (Polarisatorachse) einen Winkel $\psi = 45°$ bildet. Das Verhältnis der Feldstärkekomponenten in Richtung der „schnellen" und „langsamen" Achse ist dann 1. Nach Durchgang durch den Kristall haben beide Teilwellen eine gegenseitige Phasenverschiebung δ, so daß die linear polarisierte Welle i. allg. in eine elliptisch polarisierte Welle umgewandelt worden ist, deren große Halbachse dann parallel zur Polarisatorachse liegt. Das Achsenverhältnis b/a (Elliptizität) enthält die Information über die Phasenverschiebung δ.

Zur Bestimmung der *Elliptizität* wird die elliptisch polarisierte Welle in eine linear polarisierte Welle überführt. Das geschieht am einfachsten mit einem Kompensator, der zusätzlich so zwischen die beiden gekreuzten Polarisatoren gestellt wird, daß sein Achsenkreuz unter dem Winkel von 45° zur Schwingungsrichtung des Polarisators orientiert ist. Mit dem Kompensator wird gerade so eine Phasenverschiebung δ' eingestellt, daß die der doppelbrechenden Platte δ kompensiert wird $\delta = -\delta'$. Bei Verwendung eines Soleil-Kompensators erscheint das Gesichtsfeld des Analysators wieder dunkel.

7.5.4.4. Optische Aktivität. Eine Reihe von Stoffen (z. B. Quarz, Zucker) drehen beim Durchgang einer linear polarisierten Lichtwelle die Schwingungsebene. Derartige Stoffe werden *optisch aktiv* genannt. Der Drehwinkel α ist proportional der Länge l der durchstrahlten Probe. Erfolgt die Drehung im Uhrzeigersinn (Blickrichtung zur Lichtquelle), so wird das optisch aktive Medium *rechtsdrehend* oder *positiv* bezeichnet, bei entgegengesetzter Drehung *linksdrehend* oder *negativ*. Die Drehung der Schwingungsebene läßt sich dadurch erklären, daß die linear polarisierte Welle in dem Medium zwei zirkular polarisierte Wellen mit entgegengesetztem Umlaufsinn anregt, die sich mit unterschiedlicher Geschwindigkeit ausbreiten. Nach Durchlaufen der Probe ergibt die Überlagerung der beiden zirkular polarisierten Wellen wieder eine linear polarisierte Welle, deren Schwingungsebene aber aufgrund der Phasendifferenz, die die beiden zirkularen Wellen jetzt aufweisen, um den Winkel α gedreht worden ist. Dieser Vorgang kann als *zirkulare Doppelbrechung* aufgefaßt werden. Beschreibt man die unterschiedlichen Ausbreitungsgeschwindigkeiten der zirkularen Wellen durch die Brechzahlen n_r und n_e, so ist ihre Differenz direkt proportional dem Drehwinkel $\Delta n = n_r - n_e = \alpha\lambda/(180\,l)$, wenn α in Winkelgrad eingesetzt wird (λ = Wellenlänge in Luft, l = Probenlänge). Es ist üblich, bei Flüssigkeiten den Drehwinkel α auf die Probenlänge $l = 1$ dm und auf die Konzentration $\bar{c}$ (Masse des aktiven Stoffes in 100 g Lösung) zu beziehen. Man erhält dann die *spezifische Drehung* $[\alpha]$ als eine Stoffkonstante $[\alpha] = 100\,\alpha/(l\bar{c})$.

Bei festen Körpern (z. B. Kristallen) bezeichnet man den Drehwinkel bei einer Probendicke $l = 1$ mm mit (α).

Meßverfahren zur Bestimmung der optischen Aktivität. Die Messung des Drehwinkels α erfolgt mit einem *Polarimeter.* Es besteht im einfachsten Fall aus einer flächenhaften Beleuchtungsquelle (meist Spektrallampe), deren Strahlung durch einen feststehenden Polarisator und einen drehbaren Analysator mit Teilkreis hindurchgeht. Die Gesichtsfeldblende hinter dem Polarisator wird mit einem Fernrohr bei visueller Beobachtung auf der Netzhaut, bei fotoelektrischem Nachweis auf der Empfängeroberfläche abgebildet. Durch Drehen des Analysators wird vor und nach Einbringen der Probe auf maximale Dunkelheit eingestellt und daraus der Drehwinkel α bestimmt. Die Einstellfehler sind bei diesem Verfahren wegen des breiten Intensitätsminimums relativ groß. Verbesserte Polarimeter sind als *Halbschattengeräte* aufgebaut. Dazu wird in dem Strahlengang zwischen Polarisator und Gesichtsfeldblende zusätzlich ein kleines Polarisationsprisma (Glan-Thompson-Prisma) eingebracht, welches das halbe Gesichtsfeld einnimmt. Die Schwingungsrichtung S_2 dieses zusätzlichen Polarisationsprismas bildet mit der ursprünglichen Schwingungsrichtung S_1 einen spitzen Winkel. Die Schwingungsrichtung des Analysators kann nacheinander senkrecht zu S_1 oder S_2 eingestellt werden. Es erscheint dann die eine oder andere Gesichtsfeldhälfte dunkel. Der Differenzwinkel 2φ heißt Halbschattenwinkel (typisch 3 bis 6°). In der Mittelstellung bildet die Schwingungsrichtung des Analysators gleiche Winkel mit S_1 und S_2, so daß die Trennlinie im Gesichtsfeld verschwindet und beide Gesichtshälften gleich grau erscheinen. In dieser Einstellung ist das Polarimeter am empfindlichsten.

Messung der Rotationsdispersion. Hängt die Drehung der Schwingungsebene in einem Medium von der Wellenlänge ab, so spricht man von *Rotationsdispersion.* Wenn das Licht im blauen Teil des Spektrums stärker gedreht wird als im roten, wird sie *normale* sonst *anormale Rotationsdispersion* genannt. Die Messung der Rotationsdispersion erfolgt mit einem Polarimeter mit Halbschattenvorrichtung und einem Monochromator als wellenlängendurchstimmbare Lichtquelle.

7.6. Messung von Strahlungsgrößen (Fotometrie)

Im gesamten optischen Spektralbereich ist die optische Strahlung durch ihre *physikalischen Parameter,* wie Leistung, Energie, Wellenlänge usw., charakterisiert. Die Messung der energetischen Größen (Leistung, Leistungsdichte, Energie usw.) erfolgt mittels physikalischer Strahlungsempfänger. Zu ihrer Beschreibung werden *fotometrische* oder *strahlungsphysikalische* Größen eingeführt. Im sichtbaren Spektralbereich läßt sich die optische Strahlung durch ihren Helligkeits- und Farbeindruck beschreiben, der physiologisch vom menschlichen Auge wahrgenommen wird. Dieser „Licht"-Eindruck hängt sowohl von den energetischen Größen der elektromagnetischen Strahlung als auch von der wellenlängenabhängigen Augenempfindlichkeit $V(\lambda)$ ab. Die optische Strahlung wird hierbei nach ihrer Wirkung auf das menschliche Auge bewertet. Die $V(\lambda)$-bewerteten strahlungsphysikalischen Größen werden als *lichttechnische* Größen bezeichnet.

7.6.1. Strahlungsphysikalische Größen

Jede optische Strahlung ist mit einem Energietransport verbunden. Diese Strahlungsenergie Q_e wird in Joule (1 J = 1 Ws) gemessen. Der *Strahlungsfluß* oder die *Strahlungsleistung* $\Phi_e = dQ_e/dt$ gibt die von einer Strahlungsquelle emittierte oder von einem Empfänger aufgenommene Strahlungsenergie je Zeiteinheit (SI-Einheit Watt) an. Der Strahlungsfluß erfolgt in einem Strahlenkegel mit dem Öffnungswinkel 2σ. Es ist üblich, den Strahlenkegel durch seinen *Raumwinkel* Ω zu beschreiben. Er ist definiert als das Verhältnis der Kugeloberfläche A, die der Kegel mit der Spitze am Kugelmittelpunkt aus der Kugel herausschneidet, zum Quadrat des Kugelradius r^2 ($\Omega = A/r^2$). Die Raumwinkeleinheit Ω_0 ist 1 Steradiant (sr). Sie stellt den Teil der Oberfläche einer Einheitskugel ($r = 1$) dar, der von

dem Raumwinkel ausgeschnitten wird. Der maximale Raumwinkel ist $\Omega = 4\pi\,\mathrm{sr} = 4\pi\Omega_0$. Ein Strahlenkegel mit dem Öffnungswinkel 2σ hat den Raumwinkel $\Omega = 4\pi \sin^2 \sigma\Omega_0$. Die *Strahlstärke* I_e ist der auf den Raumwinkel $\mathrm{d}\Omega$ bezogene Strahlungsfluß $\mathrm{d}\Phi_e$: $I_e = \mathrm{d}\Phi_e/\mathrm{d}\Omega$ (SI-Einheit $\mathrm{W\,sr^{-1}}$). Die *Strahldichte* L_e ist die auf die (vom Empfänger aus gesehene) Strahlerfläche $\mathrm{d}A_1 \cos\vartheta_1$ bezogene Strahlstärke, $L_e = \mathrm{d}I_e/\mathrm{d}A_1 \cos\vartheta_1$ (SI-Einheit $\mathrm{W\,m^{-2}\,sr^{-1}}$). Bei einem *Lambert-Strahler* (z.B. Schwarzer Körper) ist die Strahldichte L_e unabhängig vom Beobachtungswinkel ϑ_1. Er erscheint unter jedem Beobachtungswinkel ϑ_1 gleich hell. Der gesamte von einem Flächenelement $\mathrm{d}A_1$ emittierte Strahlungsfluß $\mathrm{d}\Phi_e$ wird *spezifische Ausstrahlung* oder *Strahlungsflußdichte* M_e genannt, $M_e = \mathrm{d}\Phi_e/\mathrm{d}A_1$ (SI-Einheit $\mathrm{W\,m^{-2}}$). In der Optik wird die Strahlungsflußdichte meist als *Intensität* I bezeichnet. Der auf eine Empfängerfläche $\mathrm{d}A_2$ unter dem Winkel ϑ_2 auffallende Strahlungsfluß $\mathrm{d}\Phi_e$ wird als *Bestrahlungsstärke* E_e bezeichnet, $E_e = \mathrm{d}\Phi_e/\mathrm{d}A_2 \cos\vartheta_2$ (SI-Einheit $\mathrm{W\,m^{-2}}$). Bei senkrechtem Einfall ist $\vartheta = 0$. Die gesamte auf die scheinbare Fläche $\mathrm{d}A_2 \cos\vartheta_2$ auftreffende Strahlungsenergie ist die *Bestrahlung* $H_e = \mathrm{d}Q_e/\mathrm{d}A_2 \cos\vartheta_2$ (SI-Einheit $\mathrm{J\,m^{-2}}$). Da die optische Strahlung unterschiedliche Wellenlängenanteile enthält, werden weiterhin *spektrale Größen* eingeführt, die den Wert der Grundgröße im Wellenlängenbereich zwischen λ und $\lambda + \mathrm{d}\lambda$ beschreiben, z. B. der *Spektrale Strahlungsfluß* $\Phi_{e\lambda} = \mathrm{d}\Phi_e/\mathrm{d}\lambda$ (SI-Einheit $\mathrm{W\,m^{-1}}$).

7.6.2. Lichttechnische Größen

Für die meßtechnische Erfassung der vom menschlichen Auge wahrgenommenen optischen Strahlung wurden gesonderte lichttechnische Größen eingeführt. Sie entsprechen den strahlungsphysikalischen Größen, sind jedoch nicht von den energetischen Parametern der Lichtstrahlung sondern von ihrem Lichtreiz auf das Auge abgeleitet.
Die lichttechnische Grundgröße ist die *Lichtstärke* (I_v). Sie entspricht der strahlungsphysikalischen Größe Strahlstärke. Für sie wurde eine gesonderte SI-Einheit, die *Candela* (cd) eingeführt, die wie folgt definiert war: „Die Candela ist die Lichtstärke, die ein schwarzer Körper der Fläche $1/600\,000\,\mathrm{m^2}$ ($= 1/60\,\mathrm{cm^2}$) bei der Erstarrungstemperatur des Platins ($T \approx 2045$ K) und bei einem Druck $101\,325\,\mathrm{N\,m^{-2}}$ senkrecht zu seiner Oberfläche ausstrahlt." Die neue Definition lautet: „Die Candela ist die in einer Richtung abgegebene Lichtstärke einer Lichtquelle, die eine monochromatische Strahlung der Frequenz $540 \cdot 10^{12}$ Hz aussendet und deren Strahlstärke in dieser Richtung (1/683)W/sr beträgt." Alle weiteren Größen werden von dieser Grundgröße abgeleitet.
Die Leuchtdichte L_v (analog Strahldichte) ist die Lichtstärke $\mathrm{d}I_v$ je Flächeneinheit $\mathrm{d}A$: $L_v = dI_v/\mathrm{d}A$ (SI-Einheit $\mathrm{cd\,m^{-2}}$). Der *Lichtstrom* Φ_v entspricht dem vom Auge wahrgenommenen Strahlungsfluß. Eine punktförmige Lichtquelle der Lichtstärke I_v sendet in den Raumwinkel $\mathrm{d}\Omega$ den Lichtstrom $\mathrm{d}\Phi_v$ aus, $\mathrm{d}\Phi_v = I_v \mathrm{d}\Omega$ (SI-Einheit cd sr). Sie trägt den Namen Lumen (lm), 1 lm = 1 cd sr. Fällt der Lichtstrom $\mathrm{d}\Phi_v$ senkrecht auf das Flächenelement $\mathrm{d}A$, so erzeugt er die Beleuchtungsstärke $E_v = \mathrm{d}\Phi_v/\mathrm{d}A$. Die SI-Einheit Lumen/m² wird Lux (lx) genannt, 1 lx = 1 lm/m². Für bestimmte Prozesse ist die Zeitdauer der Lichteinwirkung von Bedeutung. Das Zeitintegral über die Beleuchtungsstärke E_v wird Belichtung H_v genannt, $H_v = \int E_v \mathrm{d}t$ (SI-Einheit Lux · s). Es gilt $1\,\mathrm{lxs} = 1\,\mathrm{lm\,m^{-2}s}$.

Lichttechnische Stoffkennzahlen. Zur Charakterisierung der Eigenschaften von Materialien der Lichttechnik werden die folgenden dimensionslosen Stoffkennzahlen eingeführt, deren Wert $\leqq 1$ ist:

Reflexionsgrad $$\varrho = \frac{\text{reflektierter Lichtstrom}}{\text{auffallender Lichtstrom}} = \frac{\Phi_\varrho}{\Phi}$$

Absorptionsgrad $$\alpha = \frac{\text{absorbierter Lichtstrom}}{\text{auffallender Lichtstrom}} = \frac{\Phi_\alpha}{\Phi}$$

Transmissionsgrad $$\tau = \frac{\text{durchgelassener Lichtstrom}}{\text{auffallender Lichtstrom}} = \frac{\Phi_\tau}{\Phi}$$

Es gilt $\varrho + \alpha + \tau = 1$.

7.6.3. Zusammenhang zwischen strahlungsphysikalischen und lichttechnischen Größen

Die spektrale Hellempfindlichkeitskurve des menschlichen Auges wurde international aus einer Vielzahl von Messungen als Empfindlichkeitskurve $V(\lambda)$ eines „Normalbeobachters" für das hell adaptierte Auge festgelegt. Die $V(\lambda)$-Kurve ist so normiert, daß das Maximum bei $\lambda = 555\,\mathrm{nm}$ (grünes Licht) den Wert $V(555\,\mathrm{nm}) = 1$ hat. Die $V(\lambda)$-Kurve grenzt den sichtbaren Bereich des Spektrums von 380 bis 780 nm ein. Durch Festlegung der Empfindlichkeitskurve eines „Normalbeobachters" können lichttechnische Messungen unabhängig vom menschlichen Auge durchgeführt werden, sofern die spektrale Empfindlichkeit des Empfängers der $V(\lambda)$-Kurve des Auges entspricht.
Fotometrisches Strahlungsäquivalent. Bei einer monochromatischen Lichtwelle ist der Lichtstrom $\Phi_{v\lambda} \mathrm{d}\lambda$, der im Wellenlängenbereich λ und $\lambda + \mathrm{d}\lambda$ ausgestrahlt wird, proportional der spektralen Empfindlichkeit $V(\lambda)$ des Auges und dem spektralen Strahlungsfluß $\Phi_{e\lambda}$ der Lichtquelle, $\Phi_{v\lambda}\mathrm{d}\lambda = K_m V(\lambda)\, \Phi_{e\lambda} \mathrm{d}\lambda$. K_m ist ein Umrechnungfaktor, der die Verknüpfung des Lichtstromes $\Phi_v = \Phi_{v\lambda}\mathrm{d}\lambda$ (in Lumen gemessen) mit dem vom Auge bewerteten Strahlungsfluß $\Phi_e = \Phi_{e\lambda}\mathrm{d}\lambda$ (in Watt gemessen) herstellt. Der Umrechnungsfaktor K_m hat den Wert $K_m = 673$ lm/W. Er wird *fotometrisches Strahlungsäquivalent* genannt. Sein Kehrwert $M = 1/K_m = 1{,}486 \cdot 10^{-3}$ W/lm heißt *Lichtäquivalent.* Für nichtmonochromatische Lichtquellen erhält man den Zusammenhang zwischen spektralem Strahlungsfluß $\Phi_{e\lambda}$ und Lichtstrom Φ_v:

$$\Phi_v = \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \Phi_{v\lambda} \mathrm{d}\lambda = K_m \int_{380\,\mathrm{nm}}^{780\,\mathrm{nm}} \Phi_{e\lambda} V(\lambda)\, \mathrm{d}\lambda.$$

Analog gilt der Zusammenhang zwischen den anderen strahlungsphysikalischen und ihren zugeordneten lichttechnischen Größen.

7.6.4. Grundlagen fotometrischer Messungen

Fotometrisches Grundgesetz. Eine punktförmige Lichtquelle mit der Lichtstärke I bestrahlt im Abstand r eine kleine Fläche A. Die Beleuchtungsstärke E auf dieser Fläche ist $E = (I/r^2)\,\Omega_0$. Die Abnahme der Beleuchtungsstärke mit dem Quadrat der Entfernung ist eine wesentliche Aus-

sage des fotometrischen Grundgesetzes. Es ist näherungsweise auch für flächenhafte Lichtquellen gültig, sofern die Entfernung r hinreichend groß zum Radius R der Strahlerfläche ist. Für $r \geqq 10\,R$ beträgt der Fehler $\leqq 1\%$. Trifft die Strahlung unter dem Winkel ε zur Flächennormalen von A auf, so ist die Beleuchtungsstärke $E = (I/r^2)\,\Omega_0 \cos \varepsilon$.

Schwächungsmethoden der Fotometrie. Fotometrische Meßverfahren erfordern eine definierte meßbare Schwächung des Lichtstromes. Sie kann erfolgen durch Änderung der Entfernung zwischen Lichtquelle und Meßort, durch rotierende Sektoren, durch Polarisatoren, Graufilter oder Blenden. Der rotierende Sektor wird häufig in der Fotometrie angewandt. Unter Ausnutzung des Talbotschen Gesetzes wird die Leuchtdichte dadurch herabgesetzt, daß entsprechend der Rotationsgeschwindigkeit des Sektors dem Auge periodische Impulse der Dauer $\Delta t = T\alpha/180°$ angeboten werden. Die wahrgenommene Leuchtdichte L ist dann gegenüber der auffallenden Leuchtdichte L_0 geschwächt, $L = L_0 \Delta t/T = L_0 \alpha/180°$ (T = Periodendauer, α = Öffnungswinkel der Sektorblende).

Abbesches Theorem der Leuchtdichteübertragung. Bei der optischen Abbildung einer strahlenden Fläche kann die Leuchtdichte L_B ihres Bildes höchstens gleich der Leuchtdichte L_S des Strahlers sein. Bei einem Transmissionsgrad τ der Optik gilt $L_B = \tau L_S$. Wenn die Brechzahlen vor (n_S) und hinter (n_B) der Linse unterschiedlich sind, so beträgt bei vernachlässigbaren Verlusten ($\tau = 1$) die Leuchtdichte im Bild $L_B = (n_B^2/n_S^2)\,L_S$.

7.6.5. Darstellung der Grundeinheit „Candela"

Vor 1979 wurde die Grundeinheit der Lichtstärke, die Candela, durch einen speziellen schwarzen Strahler dargestellt. Er besteht aus einem schwer schmelzbaren Röhrchen (etwa 2 mm Durchmesser, 40 mm Länge), das mit Platin gefüllt ist (Abb. 7.6.5–1).

Er wird mittels eines Hochfrequenzsenders in einem Bad aus flüssigem Platin erhitzt. Bei dessen Erstarrungspunkt beträgt dann die Leuchtdichte L_0 der strahlenden Platinoberfläche im Röhrchen $L_0 = 60\ cd\,cm^{-2}$. Präzise Messungen mit diesem Strahler sind jedoch sehr kompliziert und aufwendig. Sie waren daher wenigen Physikalischen Staatsinstituten vorbehalten. Aus diesem Grunde wurde 1979 von der 16. Generalkonferenz für Gewichte und Maße eine neue Definition der Lichtstärke (vgl. 7.6.2.) für verbindlich erklärt, die auf einer direkten Messung der Strahlstärke beruht. Sie ermöglicht wegen der in den letzten Jahren verbesserten experimentellen Meßtechnik für die strahlungsphysikalischen Größen eine Erhöhung der Meßgenauigkeit der Lichtstärke. Die Darstellung der Candela ist jetzt mit einer Genauigkeit von besser als $2 \cdot 10^{-3}$ möglich. Die Strahlung gelangt über ein geeignetes optisches System und nach Begrenzung durch eine Meßblende auf einen lichtelektrischen Empfänger. Die Öffnung der Blende (Fläche A) strahlt nach dem Abbeschen Theorem mit der Leuchtdichte $L = \tau L_0$, wenn τ der Transmissionsgrad der Optik zwischen Strahler und Blende ist, bzw. mit einer Lichtstärke $I = LA = \tau L_0 A$. Die Strahlung erzeugt im lichtelektrischen Empfänger ein elektrisches Signal bestimmter Größe. Nach Substitution des Schwarzen Strahlers durch eine speziell aufgebaute Glühlampe hoher Konstanz *(Normallampe)* können die Einheiten der Lichtstärke bzw. der Leuchtdichte auf die Normallampe übertragen werden, wenn ihre Strahlung im Empfänger ein Signal gleicher Größe ergibt. Die Messungen mit dem Schwarzen Strahler sind außerordentlich aufwendig. Wegen der relativ geringen Konstanz der Lampen wird stets eine Gruppe von vorher „gealterten" Glühlampen

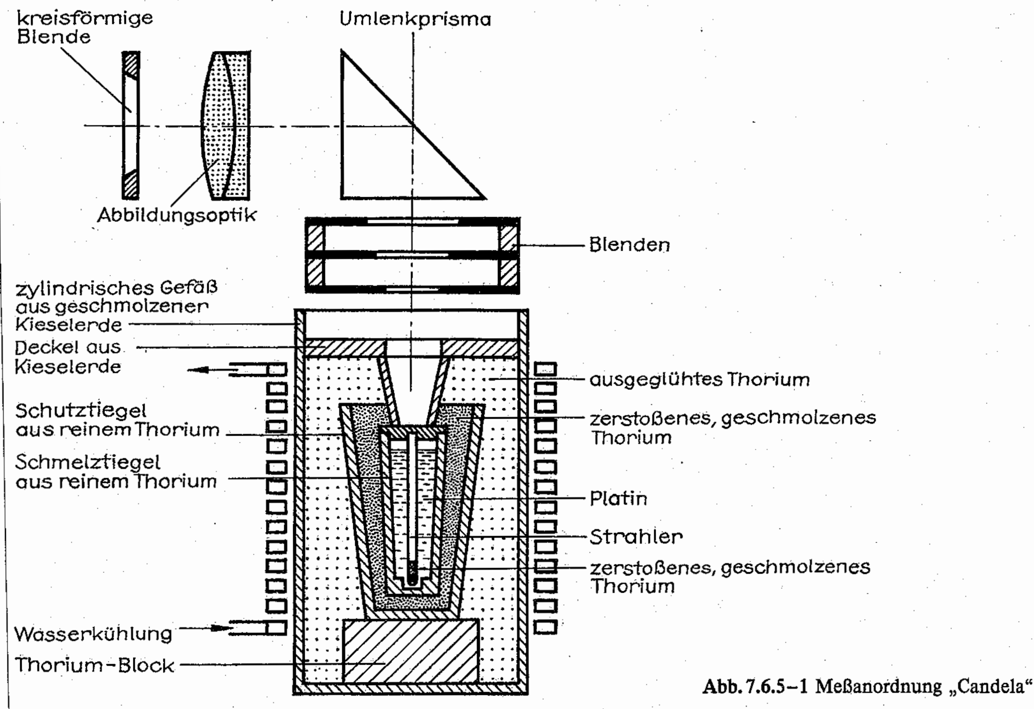


Abb. 7.6.5–1 Meßanordnung „Candela"

als *Sekundärnormale* eingemessen. Von diesen Sekundärnormalen werden seinerseits niedere Normale abgeleitet, die dann für technische Messungen verwendet werden.

7.6.6. Fotometrische Meßverfahren

Die Messung lichttechnischer Größen erfolgt mit Hilfe von *Fotometern*, welche die auszumessende Lichtquelle X direkt mit einer Normallampe vergleichen oder die mittels einer Normallampe geeicht worden sind. Bei Verwendung lichtelektrischer Empfänger (Fotoelemente, Fotowiderstände und Fotodioden auf Halbleiterbasis, Fotozellen, Sekundärelektronenvervielfacher) spricht man von *physikalischer Fotometrie*, bei der Integration des menschlichen Auges als Empfänger in das Fotometer von *visueller Fotometrie.*

Physikalische Fotometrie. Bei einem üblichen lichtelektrischen Empfänger ist in weiten Grenzen der Fotostrom proportional der auffallenden Strahlungsleistung. Als lichttechnische Größe wird daher primär der *Lichtstrom* Φ gemessen, der in die Eintrittsöffnung des Empfängers fällt. Bei dem Vergleich von Strahlungsquellen gleicher spektraler Strahlungsverteilung können lichtelektrische Empfänger beliebiger spektraler Empfindlichkeitskurven verwendet werden. Sind die spektralen Strahlungsverteilungen von Meß- und Vergleichslichtquelle unterschiedlich, so muß die relative spektrale Empfindlichkeit des Empfängers durch Farbfilter der des Auges ($V(\lambda)$-Kurve) angepaßt werden. Lichtelektrische Fotometer arbeiten nach dem Ausschlags- oder Substitutionsprinzip. Beim *Ausschlagsverfahren* wird aus der Größe des Fotostromes auf die zu messende lichtelektrische Größe geschlossen. Das Fotometer wird mit einer Normallampe geeicht. Nichtlinearitäten zwischen Fotostrom und Meßgröße werden durch das *Substitutionsverfahren* ausgeschlossen. Dabei werden Lichtquelle X und Normallampe nacheinander verglichen. Durch geeignete Schwächungsvorrichtungen werden ihre Lichtströme so eingestellt, daß die Beleuchtungsstärken auf dem Empfänger gleich groß sind. Beim *Wechsellichtverfahren* fällt die Strahlung, die vom Meßobjekt bzw. von einer Vergleichslichtquelle kommt, im periodischen Wechsel nacheinander auf den Empfänger. Durch eine Schwächungseinrichtung im Vergleichsstrahlengang werden die Lichtströme in beiden Strahlengängen so lange angeglichen, bis der Wechselstromanteil des Fotostromes zu Null wird. Fotometer nach dem Wechsellichtverfahren werden zur Bestimmung lichttechnischer Stoffkennzahlen und in der Spektralfotometrie angewendet, wo der Meß- und Vergleichsstrahlengang von der gleichen Lichtquelle abgeleitet werden kann. Dadurch lassen sich Intensitätsschwankungen der Lichtquelle ausschließen.

Visuelle Fotometrie. Da die lichttechnischen Größen von der Augenempfindlichkeitskurve abgeleitet werden, haben visuelle Meßverfahren auch heute noch ihre Bedeutung. Sie beruhen auf den Helligkeitseindrücken, die von den unterschiedlichen *Leuchtdichten* zweier Strahlungsquellen herrühren. Das Auge kann nur entscheiden, ob die Leuchtdichten im Gesichtsfeld gleich sind oder nicht (Nullabgleichmethode). Dazu werden benachbarte Flächen, die von den verschiedenen Lichtquellen beleuchtet werden, entweder auf gleiche Helligkeit *(Gleichheitsfotometer)* oder auf gleichen Kontrast *(Kontrastfotometer)* eingestellt. Diese Meßverfahren sind insbesondere geeignet, wenn Lichtquelle X und Normallampe die gleiche relative spektrale

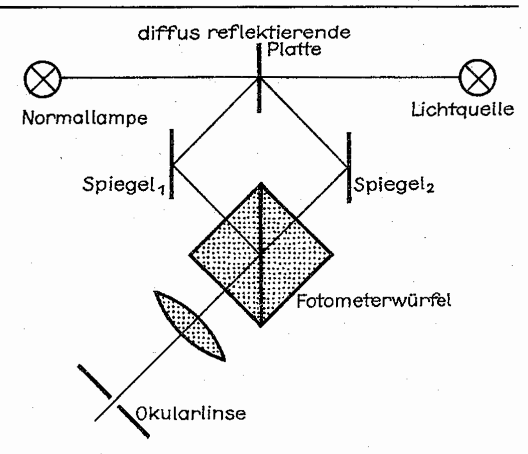


Abb. 7.6.6–1 Lummer-Brodhun-Würfel

Strahlungsverteilung aufweisen. Der Vergleich der Leuchtdichten erfolgt in speziellen *Fotometerköpfen.* Ein sehr verbreitet eingesetzter Fotometerkopf wurde von LUMMER und BRODHUN angegeben. Eine weiße, diffus reflektierende Platte wird auf der einen Seite von der Lichtquelle X und auf der anderen Seite von der Normallampe beleuchtet. Das diffus reflektierte Licht gelangt über zwei Spiegel in den Fotometerwürfel, der aus zwei 90°-Prismen besteht, die mit ihrer Diagonalseite zusammengesetzt sind (Abb. 7.6.6–1). Ein Teil der Fläche hat optischen Kontakt,

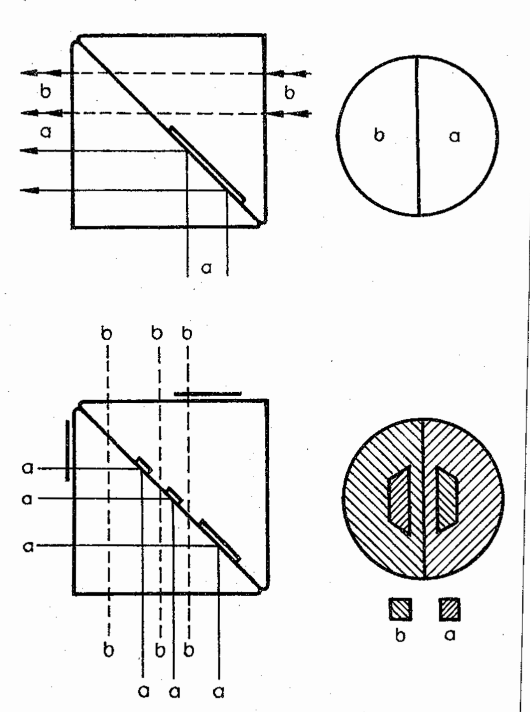


Abb. 7.6.6–2 Fotometerwürfel, *links* Schnitt, *rechts* Gesichtsfeld

so daß dort die Strahlung hindurchgeht, andere Teile wurden herausgeätzt, was zu einer Totalreflexion der Strahlung in diesen Bereichen führt. Dadurch liegen im Gesichtsfeld Bereiche nebeneinander, die von unterschiedlichen Lichtquellen beleuchtet werden. Die Einstellung der Beleuchtungsstärken bei einer Aufteilung des Gesichtsfeldes in zwei Hälften (Abb. 7.6.6–2) erfolgt so, daß beide Felder gleiche Helligkeit aufweisen und die Trennlinie verschwindet *(Gleichheitsfotometer)*. Eine modernere Ausführung hat im Inneren der beiden Hälften zusätzliche Trapezfelder, die jeweils von einer Lichtquelle ausgeleuchtet werden, während die andere Lichtquelle das Umgebungsfeld beleuchtet. Die Strahlung für die Trapezfelder wird noch durch zusätzliche Kontrastplättchen geschwächt. Es wird jetzt auf gleichen Kontrast zwischen Trapezfeld und Umgebung in beiden Hälften eingestellt *(Kontrastfotometer)*. Die Einstellgenauigkeit beträgt bei gleichfarbigen Lampen ≈ 0,5 bis 1%. Sie nimmt bei verschiedenfarbigen Lampen zu. Für die Messung verschiedenfarbiger Lampen ist das *Flimmerprinzip* günstiger. Hierbei wird das Licht dem Auge im schnellen Wechsel nacheinander angeboten. Bei einer geringen Wechselfrequenz tritt zunächst ein farbiges Flimmern auf, das oberhalb einer Grenzfrequenz verschwindet. Es bleibt ein Flimmern der Helligkeit übrig. Dieses Helligkeitsflimmern erreicht ein Minimum, wenn beide Leuchtdichten gleich sind.

7.6.7. Messung spezieller lichttechnischer Größen

Leuchtdichte L wird bei visuellen Verfahren direkt gemessen, sofern von der strahlenden Fläche A_s die Augenpupille voll ausgeleuchtet wird. Da der Fotostrom lichtelektrischer Empfänger proportional dem Lichtstrom Φ ist, ergibt sich für die Leuchtdichte L bei senkrechtem Einfall $L = (\Phi/A_E)\,\Omega$ (A_E = Empfängerfläche, Ω = Raumwinkel, unter dem vom Empfänger aus der Strahler erscheint).

Lichtstärke I wird immer über die von ihr erzeugte Beleuchtungsstärke E gemessen. Für einen Lichteinfall unter dem Winkel ε zur Flächennormalen der beleuchteten Fläche gilt $E = (I/r^2)\,\Omega_0 \cos\varepsilon$.

Beleuchtungsstärke E. Der Empfänger des visuellen oder physikalischen Fotometers wird mit seiner Auffangfläche an die Stelle gebracht, wo die Beleuchtungsstärke gemessen werden soll. Geräte zur Messung der Beleuchtungsstärke („Luxmeter") sind heute in der Regel mit lichtelektrischen Empfängern ausgerüstet. Die Beleuchtungsstärke E ergibt sich aus dem Lichtstrom, der auf die Empfängerfläche A_E fällt, $E = \Phi/A_E$.

Lichtstrom Φ. Eine einfache und schnelle Messung des gesamten Lichtstromes einer Lichtquelle kann mit dem *Kugelfotometer* erfolgen. Es besteht aus einer innen mattweiß gestrichenen Hohlkugel von 1 bis 3 m Durchmesser mit einer kleinen lichtdurchlässigen Öffnung, dem Meßfenster *(Ulbrichtsche Kugel)*. Die auszumessende Lichtquelle ist im Inneren der Kugel aufgehängt. Durch einen Schirm wird verhindert, daß direktes Licht auf das Meßfenster fällt. Die Beleuchtungsstärke E auf dem Meßfenster ist dann proportional dem Gesamtlichtstrom Φ, $E = \dfrac{\Phi}{4\pi r^2}\,\dfrac{\varrho}{1-\varrho}$, wobei r der Kugelradius und ϱ der Reflexionsgrad ist. Sie ist unabhängig von der Ausstrahlungscharakteristik und dem Aufhängungsort der Lampe. Der Lichtstrom wird vorwiegend nach der *Substitutionsmethode* gemessen. Dazu werden nacheinander die zu messende Lichtquelle X und eine Normallampe mit dem bekannten Lichtstrom Φ_N an die gleiche Stelle der Kugel gehängt und das Verhältnis der Beleuchtungsstärken E_X/E_N am Meßfenster gemessen. Der gesuchte Lichtstrom ist dann $\Phi_X = \Phi_N E_X/E_N$.

7.7. Elektronenmikroskopie

7.7.1. Grundlagen der Elektronenmikroskopie

Fortschritte in der Anwendung elektronenmikroskopischer Methoden gestatten seit einigen Jahren, Abbilder von Strukturen sehr dünner kristallisierter Festkörper mit nahezu atomarer Auflösung zu erhalten. Nach der Konstruktion magnetischer Elektronenlinsen durch BUSCH, die eine Brechung der Elektronenstrahlen analog den der Lichtstrahlen erzeugten, wurden maßgebliche Impulse zur Entwicklung von Durchstrahlungselektronenmikroskopen (Transmissions-Elektronenmikroskopie, TEM) etwa um 1930 von KNOLL, RÜDENBERG, RUSKA, VON BORRIES und MARTON ausgelöst. Wenig später führte M. VON ARDENNE die Methode des Rasterelektronenmikroskops (Scanning-Elektronenmikroskopie, SEM) zum Studium von Oberflächen ein. Kommerziell sind die letztgenannten Geräte seit den 60er Jahren verfügbar. Sie haben in der Anwendungsbreite die Durchstrahlungselektronenmikroskope bereits überflügelt.

7.7.1.1. Wechselwirkung von Elektronen mit Festkörpern. Zur Bilderzeugung bzw. zur Informationsgewinnung einer bestrahlten Probe können prinzipiell alle bei der Wechselwirkung von Elektronen mit Materie auftretenden Erscheinungen verwendet werden (Abb. 7.7.1–1). Die zur Analyse benutzten Elektronenstrahlen bieten dank ihrer ausgezeichneten Fokussierbarkeit bereits heute alle Möglichkeiten zur Analyse kleinster Mikrobereiche. Die moderne Elektronenmikroskopie ist inzwischen untrennbarer Bestandteil des weiten Gebietes der Festkörperanalyse geworden.

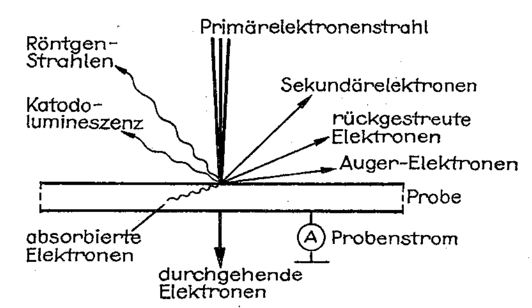


Abb. 7.7.1–1 Wechselwirkung von Elektronen mit Festkörpern, die zur Bilderzeugung in der Elektronenmikroskopie angewendet werden

7.7.1.2. Auflösungsvermögen. Die Auflösung wird gewöhnlich als kürzeste Entfernung zwischen zwei Punkten oder zwei Linien definiert, die man als unterschiedliche Bilder wahrnehmen kann. Bei stets vorhandenen Beugungserscheinungen ist im Fall des optischen Mikroskopes die Auflösung d infolge der sphärischen und chromatischen Abberation durch die Wellenlänge der zur Abbildung verwendeten Strahlung gemäß $d = 0{,}61\,\lambda/A = 0{,}61\,\lambda/(n \sin\alpha)$ auf einige Zehntel Mikrometer begrenzt. λ = Wellenlänge des verwendeten Lichtes, A = numerische Apertur nach ABBE, n =

Brechungsindex und α = halber Öffnungs- oder Aperturwinkel.
Es ist deshalb zu erwarten, daß kurzwelligere Strahlung, z. B. Röntgen- oder γ-Strahlen, bessere Auflösungsbedingungen liefern, wenn für diese Strahlungen brechende Medien existieren. Für Elektronen ist das tatsächlich der Fall. Gemäß der Beziehung von DE BROGLIE kann einem sich mit der Geschwindigkeit v bewegten Teilchen der Masse m eine Wellenlänge $\lambda = h/(mv)$ zugeordnet werden. Die Größe $h = 6{,}63 \cdot 10^{-34}$ Js stellt die Plancksche Konstante dar. Damit ergeben sich die in Tab. 7.7.1–2 für einige ausgewählte Beschleunigungsspannungen aufgeführte Wellenlängen für Elektronen, wie sie in kommerziellen Elektronenmikroskopen möglich sind.

Tab. 7.7.1–2 Wellenlänge beschleunigter Elektronen

Spannung U in kV	80	100	120	200	500	1 000
Wellenlänge λ in nm	0,004	0,003 7	0,003 3	0,002 5	0,001 4	0,000 8

Die Werte für die erzielbare Auflösung bei gegebener Wellenlänge werden nicht erreicht, da die Korrekturmöglichkeit für Abbildungsfehler in der Elektronenoptik wesentlich geringer als in der Lichtoptik ist. Insbesondere der Öffnungsfehler der Objektivlinse beschränkt die theoretische Auflösungsgrenze. SCHERZER berechnete die minimale Auflösung $d_{min} = 0{,}43\,(\lambda^3 C_\ddot{o})^{1/4}$, $C_\ddot{o}$ = Öffnungsfehlerkonstante der Objektivlinse. Der Wert für $C_\ddot{o}$ beträgt bei Hochleistungsobjektiven ≈ 1 mm. Daraus läßt sich für Beschleunigungsspannungen von 100 kV eine theoretische Auflösungsgrenze von $\approx 0{,}2$ nm erreichen, wenn eine optimale Objektivapertur $\alpha_{opt.} \approx 10^{-2}$ eingehalten wird. Diese Werte werden von modernen mechanisch und elektrisch hochstabilisierten Hochauflösungselektronenmikroskopen heute praktisch erreicht. Die dazugehörige förderliche Vergrößerung M kann man über die Auflösungsgrenze des menschlichen Auges von $d_1 \approx 0{,}1$ mm über $M = d_1/d_{min}$ errechnen und erhält Werte von ≈ 1 Million.

7.7.1.3. Elektronenmikroskopische Untersuchungsverfahren. In der Tab. 7.7.1–3 sind die gegenwärtig gebräuchlichsten elektronenoptischen Methoden systematisiert. Es hat sich bewährt, die Elektronenmikroskopie in die Durchstrahlungsmikroskopie, die elektronenoptische Direktabbildung von Oberflächen und die analytische Elektronenmikroskopie einzuteilen. Bei der letzteren Methode findet eine Kombination von elektronenoptischer Abbildung und Elektronenspektroskopie statt, bei der die infolge unelastischer Wechselwirkungsprozesse emittierte Röntgenstrahlung bzw. die Energieverluste transmittierter Elektronen zum Elementnachweis ausgenutzt werden. Moderne Geräte erlauben eine Erweiterung des Grundgerätetyps, so daß die in der Tab. 7.7.1–3 aufgeführten Möglichkeiten kombiniert in einer kommerziellen Einrichtung realisiert werden können. Im weiteren soll von der Einteilung nach Tab. 7.7.1–3 ausgegangen und einige Gesichtspunkte näher erläutert werden.

7.7.2. Durchstrahlungselektronenmikroskopie

7.7.2.1. Experimentiertechnik. Elektronenmikroskope mit Elektronenenergien von 80 bis 1000 keV können je nach der zu untersuchenden Substanz Objekte einer Dicke von höchstens einigen Zehntel Mikrometern durchdringen. Für die meisten zu untersuchenden Objekte heißt das, sie müssen einer geeigneten Präparation unterzogen werden. Üblicherweise setzt man die chemische, elektrochemische oder die Ionenstrahlabdünnung bei physikalisch-technischen Objekten ein, während im medizinisch-biologischen Bereich z. B. die Ultramikrotomdünnschnittechnik Anwendung findet, die Dünnschnitte bis ≈ 10 nm Dicke erlaubt.
Die Strahlengänge eines Elektronenmikroskops zeigt Abb. 7.7.2–1. Zwischen dem Abbildungsstrahlengang im Lichtmikroskop und im Elektronenmikroskop existieren sehr enge Analogien, wobei der Lichtquelle die Elektronenkanone und den Glaslinsen die erwähnten Elektronenlinsen entsprechen. In der Elektronenmikroskopie werden fast ausschließlich magnetische Elektronenlinsen verwendet, die rotationssymmetrische Magnetfelder zur geeigneten Ablenkung der Elektronen liefern. Im oberen Teil der Mikroskopsäule werden die Elektronen erzeugt, die für eine ausreichende Beleuchtungsintensität sorgen. Neben haarnadelförmig gebogenen Wolframkatoden haben für höhere Richtstrahlwerte, d. h. höhere Helligkeit LaB_6- oder Feld-

Tab. 7.7.1–3 Übersicht über elektronenmikroskopische Methoden

Durchstrahlungselektronenmikroskopie			elektronenmikroskopische Direktabbildung von Oberflächen	analytische Elektronenmikroskopie
Beugungsverfahren	Durchstrahlungsabbildung	indirekte Oberflächenabbildung		
	amorphe Objekte kristalline Objekte (Molekül- und Atomabbildung, Netzebenenabbildung, Abbildung von Kristalldefekten im Beugungskontrast)	Abdrucktechniken Oberflächendekoration	Oberflächenrasterelektronenmikroskopie in verschiedenen Abbildungsmoden Emissionsmikroskopie Spiegelelektronenmikroskopie	energiedispersive Mikroanalyse wellenlängendispersive Mikroanalyse Elektronenenergieverlustanalyse

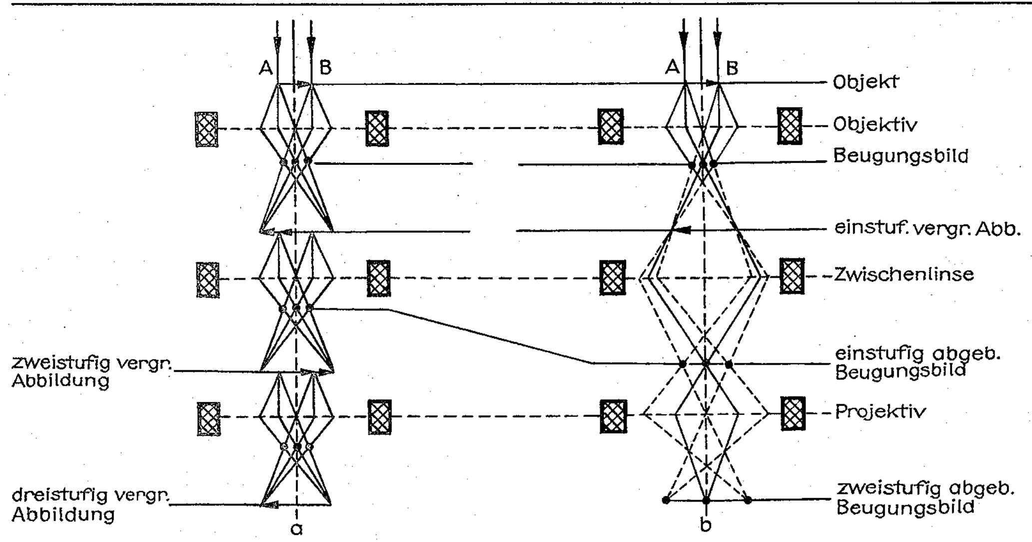


Abb. 7.7.2-1 Strahlengänge für (a) Abbildung und (b) Elektronenbeugung in einem dreistufigen Elektronenmikroskop

emissionskatoden eine entscheidende Verbesserung gebracht. Zur Regulierung der Emission ist die Katode von einem Wehnelt-Zylinder umgeben, der auf ein negatives Potential von ≈ -100 V gelegt wird. Durch unterhalb der Anode liegende zweistufige Kondensorlinsen werden die beschleunigten Elektronen zur kohärenten Beleuchtung auf die Probe gelenkt. Nach der Wechselwirkung mit dieser durchlaufen sie ein elektronenoptisches Vergrößerungssystem, bestehend aus Objektiv-, Zwischen- und Projektivlinse, und man erhält auf einem Leuchtschirm ein vergrößertes Abbild des Objektes. Eine Registrierung ist über Fotoplatte, Film, Fernseh- oder Filmkamera möglich.

Zur Vermeidung unerwünschter Streuungen des Elektronenstrahles ist es notwendig, im Vakuum zu arbeiten und zur Erzeugung sehr kleiner Wellenlängen eine Hochspannung zur Beschleunigung der Elektronen zu verwenden. Das bedingt einen hohen apparativen Aufwand im Gegensatz zum Lichtmikroskop. Ein besonderer Vorteil des Elektronenmikroskops besteht darin, daß eine Kombination von Abbildungs- und Beugungsmethoden in modernen Geräten sehr einfach realisiert ist. Einmal erzeugt man durch die Wirkung der Objektivlinse eine Abbildung durch Fokussierung der von den Objektpunkten ausgehenden Elektronen in zugeordneten Bildpunkten, andererseits werden in der hinteren Brennebene des Objektivs jeweils in die gleiche Richtung gebeugte Elektronen in einem Punkt vereinigt. Je nach der Natur des Objektes entstehen z. B. bei einkristallinen Proben Punktreflexe oder bei polykristallinen Proben die bekannten Ringdiagramme. Damit sind Aussagen zur Gitterstruktur des Festkörpers und seiner Morphologie aus identischen Gebieten erhältlich.

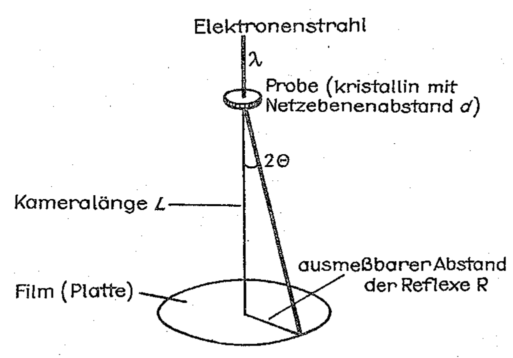


Abb. 7.7.2-2 Veranschaulichung der geometrischen Verhältnisse zur Auswertung von Elektronenbeugungsaufnahmen

7.7.2.2. Elektronenbeugung. Der Grundprozeß für die Beugung wie für die Bildentstehung ist die Streuung der Elektronen an den Atomen innerhalb der Probe. Im Gegensatz zu den Röntgenstrahlen werden die Elektronen hauptsächlich durch die Atomkerne gestreut. Für die Elektronenbeugung ist diese elastische Streuung maßgebend. Ein geringer Teil der Elektronen erleidet durch Wechselwirkung mit den Atomelektronen eine Richtungs- oder Geschwindigkeitsänderung. Dieser unelastische Anteil wird in der analytischen Elektronenmikroskopie insbesondere zur Elementidentifizierung ausgenutzt.

Bei Bestrahlung kristalliner Proben existiert innerhalb der Bausteine eine konstante Phasenbeziehung zwischen den Wellen, und man erhält kohärente elastische Streuung in eine Richtung, die als Bragg-Reflexion bekannt ist. In der Abb. 7.7.2-2 sind die geometrischen Verhältnisse veranschaulicht. Falls die Wegdifferenz $2d \sin \Theta = n\lambda$ beträgt, werden die an den Netzebenen reflektierten Wellen in Phase sein, d = Netzebenenabstand, Θ = Glanzwinkel, λ = Wellenlänge, n = Ordnung der Interferenz.

Durch die Braggsche Reflexionsbedingung ist die Richtung, in der ein Beugungsmaxima zu erwarten ist, eindeutig vorgegeben, und umgekehrt läßt sich der Netzebenenabstand exakt aus den Messungen des Beugungswinkel 2Θ ermitteln.

Elektronen haben nur Vorwärtsstreuung in einem sehr schmalen Winkelbereich. Betrachtet man bei Kupfer mit einer Gitterkonstanten $a = 0{,}351$ nm den (420-)Reflex und berechnet den Winkel zwischen Primär- und abgebeugtem Strahl, so ergibt sich für Cu-K_α-Röntgenstrahlung der Wel-

lenlänge $\lambda = 0{,}154$ nm ein Winkel $2\Theta = 158°$, während der Winkel für Elektronenstrahlen von 80 kV (entsprechend einer Wellenlänge $\lambda = 0{,}0042$ nm) $2\Theta = 1{,}5°$ beträgt. Abb. 7.7.2-2 vermittelt vereinfacht die geometrischen Relationen für die Beugungserscheinung an Objekten im Elektronenmikroskop, aus denen sich $R : d = \lambda L$ ableiten läßt. R stellt den ausmeßbaren Abstand der Reflexe auf der Fotoplatte dar und λL die sogenannte Kamerakonstante. Die letzte Größe läßt sich experimentell über eine Eichsubstanz ermitteln, z. B. Thalliumbromid, so daß die Netzebenenabstände unbekannter Substanzen auswertbar sind. Aus der Beziehung ist ersichtlich, daß man auf der Fotoplatte einen Ausschnitt des sogenannten reziproken Gitters abbildet, aus dem dann auf das Realgitter geschlossen werden kann.

Die dargelegte einfache geometrische Theorie erlaubt bereits die Lösung vieler physikalischer und technischer Aufgabenstellungen. Sie läßt jedoch keine Aussagen über die Intensitäten der einzelnen Reflexe zu. Dazu muß man die kinematische oder als beste Näherung die dynamische Theorie für Beugungs- und Kontrasterscheinungen anwenden.

Die kinematische Theorie der Elektronenbeugung geht von der Streuung der Elektronenwellen an Gitterbausteinen aus. Die Gesamtstreuintensität ergibt sich durch Superposition aller Einzelstreuwellen. Ausgangspunkt ist die Schrödinger-Gleichung für die Wellenfunktion des im Potentialfeld eines Atoms gestreuten Elektrons. Die Lösung der Gleichung erfolgt nach der Störungstheorie. Die Verwendung der ersten Bornschen Näherung hat Nachteile, da sie nur die einfach gestreute Welle berücksichtigt. Diese Näherung wird mit zunehmender Ordnungszahl schlechter. Hier muß man von höheren Näherungen ausgehen. Diese liefern keine reellen Atomformamplituden wie die Bornsche Näherung, sondern nur komplexe Größen. Physikalisch bedeutet das, es findet eine Phasenänderung statt. Deshalb ist die Gültigkeit der kinematischen Theorie auf die Streuung mit geringer Phasendifferenz sowie auf die Streuung an sehr dünnen Kristallen beschränkt, wo die Wechselwirkung zwischen reflektierten Strahlen und Primärstrahl vernachlässigt werden kann. Mit der Anwendung der dynamischen Theorie fallen die Einschränkungen weg.

Die Berücksichtigung der Phasenänderung bei der Streuung hat zur Folge, daß i. allg. die Friedelsche Regel nicht erfüllt ist, so daß enantiomorphe Kristalle in diesem Fall unterschieden werden können. Analoge Aussagen gelten bei Röntgenfeinstrukturuntersuchungen.

Die Kristallstrukturuntersuchungen mit der Elektronenbeugung lassen sich im wesentlichen nach zwei Gesichtspunkten einteilen:

1. Bestimmung des Kristallsystems, der Translationsgruppe und der Gitterdimensionen. Für die Ermittlung des Kristallsystems und der Gitterdimensionen reichen Lagemessungen der Reflexe aus, und eine Bestimmung ist bei hochsymmetrischen Kristallen über alle Arten von Beugungsdiagrammen einschließlich Debye-Scherrer-Aufnahmen möglich. Bei niedrigsymmetrischen Kristallen, wie monoklinen und triklinen, mit Ausnahme des Debye-Scherrer-Ringdiagramms. Die Translationsgruppe kann ebenfalls nicht aus dem Ringdiagramm entnommen werden.
2. Bestimmung der Raumgruppe und der Atomlagen. Der Strukturfaktor der kinematischen Theorie bzw. die Fourier-Koeffizienten der dynamischen Theorie können aus allen Beugungsdiagrammen durch Intensitätsmessungen ermittelt werden.

In der Praxis ist die Bestimmung der Raumgruppe und der Atomlagen auf die Fälle beschränkt, in denen dynamische Effekte vernachlässigbar sind. Analog zur Röntgenfeinstrukturanalyse kann man nur die Intensitäten messen, also die Beträge des Strukturfaktors. Eine genäherte Raumgruppenbestimmung ist über die Auslöschungen erreichbar. Eine vollständige Kristallstrukturbestimmung ist durch Fourier-Analyse möglich, wenn die kinematische Mehrfachstreuung und Abweichungen von der Bornschen Näherung berücksichtigt werden.

Die Größe der im Elektronenmikroskop zu untersuchenden Kristalle muß einerseits unter 100 nm liegen, wenn die kinematische Theorie anwendbar sein soll, andererseits möchte sie auch mindestens 10 nm sein, um Einflüsse von Kristallgröße und -form auf das Beugungsbild auszuschalten. Aus diesem Grund wird die Elektronenbeugung für die Kristallstrukturanalyse seltener herangezogen als die Röntgendiffraktometrie, bei der keine solchen engen Grenzen für die Kristallabmessungen existieren. Die elastische Streuung von Elektronen an den Kernen führt zu starken Beugungsintensitäten, die um den Faktor 10^6 bis 10^8 die Röntgenintensitäten übersteigen. Deshalb gestattet es die Elektronenbeugung, extrem geringe Substanzmengen zu analysieren. So kann man mit Feinbereichsbeugungsmethoden heute $\approx 10^{-15}$ g und mit speziellen Mikrobeugungsverfahren $\approx 10^{-19}$ g noch nachweisen. Durch Verwendung des Rasterfeinstrahles lassen sich in naher Zukunft Informationen aus Bereichen von wenigen Nanometern erhalten. Die Grenze für die Nachweisempfindlichkeit von Substanzen wird in der hochauflösenden Elektronenmikroskopie bei der Abbildung atomarer und molekularer Strukturen mit $\approx 10^{-22}$ g bezogen auf Atom- oder Molekularcluster erreicht.

7.7.2.3. Elektronenmikroskopische Bildkontraste. Zur Feststellung der Konfiguration eines Objektes ist ein Bildkontrast erforderlich. Er bestimmt die tatsächliche Auflösung eines Elektronenmikroskops. Sieht man von Einflüssen des elektronenoptischen Systems ab, so wird der Bildkontrast hauptsächlich durch Streuung bzw. Beugung der abbildenden Elektronen am Objekt erzeugt. In der Elektronenmikroskopie handelt es sich stets um einen Phasenkontrast im weitesten Sinne. Man unterscheidet innerhalb dessen den Streuabsorptionskontrast vom Beugungskontrast und dem Phasenkontrast im engeren Sinne. Für amorphe Objekte wird der Bildkontrast in der Durchstrahlungsabbildung durch die Massendicke bestimmt und als Streuabsorptionskontrast bezeichnet. Die Intensität I des ungestört einen Objektbereich durchlaufenden Elektronenbündels hängt mit der Intensität des auf diesen Objektbereich auffallenden Strahlenbündels I_0 über die Beziehung $I = I_0 \exp(-\sigma\varrho z)$ zusammen, wobei σ den Streuquerschnitt, ϱ die Dichte des Objektes und z die Dicke des durchstrahlten Bereiches bzw. ϱz die Massendicke bezeichnen. Der Kontrast selbst ist durch die Relation $K = (I_0 - I)/I_0 = 1 - I/I_0$ definiert. Werte zwischen 0,05 und 0,1 sind fotografisch noch nachweisbar.

Das Hauptanwendungsgebiet des Streuabsorptionskontrastes liegt in der Medizin und der Biologie. Er hat aber auch für die ein- oder mehrstufige indirekte Abbildung von Oberflächen mittels der Abdrucktechnik Bedeutung. Hier erfolgt eine Kohle-Platin-Schrägbeschattung der eigentlichen Oberfläche im Vakuum. Der geschlossene Abdruckfilm wird mit geeigneten Mitteln abgelöst und vermittelt in der Durchstrahlungsabbildung einen plastischen Eindruck der ursprünglichen Oberfläche. Damit sind gegebenenfalls

Stufenhöhen von wenigen Nanometern abbildbar. Atomare Stufenhöhen lassen sich damit nicht erfassen. Für diese Zwecke eignet sich die Methode der Oberflächendekoration.

Der Kontrast kristalliner Proben entsteht hauptsächlich durch die Bragg-Reflexionen. Er ist stark von der Orientierung der Probe zum Elektronenstrahl abhängig. Durch Interferenz des ungebeugten Strahles mit gebeugten Elektronenbündeln ist eine Abbildung ausgewählter Netzebenen im Kristall möglich. Dazu nutzt man die unkomplizierte Umschaltung zwischen einem exakt bekannten Beugungsbild, das eine genaue Ausrichtung der Probe mit Hilfe eines eingebauten Goniometers zum Elektronenstrahl erfordert, und der Abbildung des Objektes. Mit Hilfe einer Objektivblende bekannten Durchmessers werden die Reflexe ausgewählt, die zum Bild beitragen sollen. Im Abbildungsmodus wird dann die Vergrößerung eingestellt, um die gewünschten Netzebenen sichtbar zu machen. Die abzubildende Struktur wird bei einer bestimmten Unterfokussierung des Bildes beobachtbar. Es ist ein stark defokussierungsabhängiger Phasenkontrast im engeren Sinne. Diese Art der Abbildung stellt allerdings hohe Forderungen an die Justierung des Gerätes und an die Probenpräparation. Es läßt sich zeigen, daß hierbei besonders günstige Voraussetzungen für eine Abbildung bestehen. Man erhält eine Auflösung von ≈ 0,1 nm, die unter der theoretisch angegebenen Punktauflösung liegt. Die Methode der Netzebenenabbildung wird heute allgemein auch zum Auflösungstest eines Elektronenmikroskops verwendet. Existierende Kristalldefekte können dabei direkt durch die Störung im Abbild der Netzebenen sichtbar werden.

Der gegenwärtige Stand der elektronenmikroskopischen Gerätetechnik mit Hochleistungsobjektiven erlaubt mit einer Auflösung von ≈ 0,2 nm die Abbildung von Atomen. Schwierigkeiten bereiten der erforderliche Bildkontrast, der die Untersuchungen auf schwerere Atome beschränkt, und die durch den Elektronenstrahl hervorgerufene Objektschädigung.

Die elektronenmikroskopische Atomabbildung von Uranium und Thorium gelang Crewe 1970 mit einem Rasterelektronenmikroskop. Ein Jahr später konnte Hashimoto mit einem konventionellen Transmissionselektronenmikroskop (TEM) die Atomabbildung verwirklichen, besser ausgedrückt, das Bild identischer Atome im Kristall, die in einer Richtung entlang der optischen Achse der Linse gestapelt sind, erzeugen. Molekülabbildungen sind ebenfalls möglich. Allerdings erwiesen sich hier die Fragen der Bildinterpretation als schwieriger, und eindeutige Aussagen bei der Abbildung molekularer Strukturen können nur erhalten werden, wenn mit der experimentellen elektronenmikroskopischen Aufnahme der Abbildungsprozeß gleichzeitig rechentechnisch erfaßt werden kann. Das erste Beispiel für eine Molekülabbildung gelang Uyeda 1970 von Hexadecachlor-Cu-Phthalocyanin-($Cl_{16}CuPc$)-Aufdampffilmen.

Diese Mustersubstanz für die molekulare Abbildung hat eine ≈ 40fach höhere Strahlenbelastbarkeit als die im Prinzip schon relativ stabilen CuPc-Filme und übertrifft damit andere organische Moleküle wesentlich. Ausführliche Studien des Strahlenschädigungsmechanismus stammen von Uyeda, Kobayashi und Reimer bzw. von Murata. Einen weiteren Fortschritt in der Qualität der Molekülabbildung brachte die Anwendung der Höchstspannungselektronenmikroskopie, da die Wellenlänge die Auflösung stark beeinflußt und einzelne Linsenfehler relativ klein gehalten werden können.

Für die numerische Berechnung von Strukturabbildungen kann die kinematische Näherung von Uyeda und Ishizuka verwendet werden. Aus den Strukturfaktoren der kristallinen Probe wird ein Beugungsdiagramm konstruiert, das nach Multiplikation mit der Kontrastübertragungsfunktion durch eine inverse Fourier-Transformation die Abbildungswellenfunktion ergibt. Die berechnete Bildintensität kann ausgedruckt und mit der experimentellen Aufnahme verglichen werden. Gleichzeitig findet die optische Beugung Anwendung, mit der die Abbildungsbedingungen sowie Bildfehler bei der routinemäßigen Analyse von Hochauflösungsaufnahmen kontrolliert werden können.

Bei der Abbildung von Kristalldefekten im Durchstrahlungselektronenmikroskop greift man im wesentlichen auf das von Hirsch, Horne und Whelan bzw. von Bollmann eingeführte Verfahren des Beugungskontrastes zurück. Bei dieser Methode werden nur die von der Gitterperiodizität abweichenden Kristallstörungen erfaßt und nicht die einzelnen Netzebenen aufgelöst. Damit sind die geräte- und präparationstechnischen Anforderungen nicht so hoch, wie bei der von Menter vorgeschlagenen Netzebenenabbildung. Die Abb. 7.7.2–3 zeigt den Hellfeld-Strahlengang im Beugungskontrast. Die in der hinteren Brennebene des Objektivs vorhandene Kontrastblende ist so angeordnet, daß nur das ungebeugte Strahlenbündel zur Bildebene kommt. Als Bildkontrast eines idealen ungestörten Gitters erhält man eine gleichmäßig ausgeleuchtete Bildfläche. Nur in Bereichen, wo der Kristall lokale Abweichungen z. B. durch ein- oder mehrdimensionale Baufehler zeigt, ergeben sich gegenüber der Umgebung geänderte Bragg-Orientierungen.

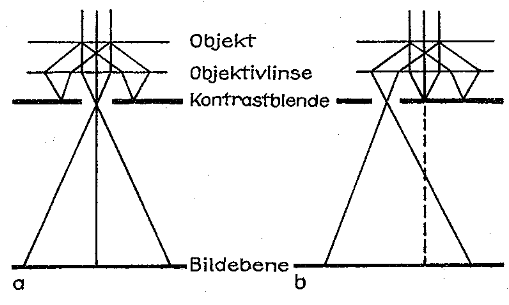


Abb. 7.7.2–3 Strahlengang für Untersuchungen im Beugungskontrast, (a) Hellfeld, (b) Dunkelfeld

Diese Bereiche werden aufgrund ihres abweichenden Beugungsvermögens von der ungestörten Umgebung sichtbar. Verständlich wird diese Erscheinung, wenn man die dynamische Theorie der Elektronenbeugung in der Näherung des Zweistrahlfalles verwendet, d. h. neben dem Nullreflex wird nur ein weiterer starker Beugungsreflex verwendet. Die Intensität des gebeugten Elektronenbündels nach Durchlaufen eines von Kristallbaufehlern freien Stückes der Dicke t wird durch die Beziehung $I \sim \sin^2 \pi t s' / (\pi s')^2$ gegeben, mit $s' = (1/\zeta_g^2 + s^2)^{1/2}$. Die Größe s bezeichnet den Anregungsfehler, d. h. die Abweichung des zur Abbildung verwendeten reziproken Gittervektors g von der idealen Bragg-Lage, und ζ_g die Extinktionslänge, eine materialabhängige Größe mit Werten von einigen 10 nm. Man erhält bei konstantem Anregungsfehler infolge einer Änderung der Kristalldicke t eine periodische Intensitätsmodulation, die physikalisch über die Ewaldsche Pendellösung zu interpretieren ist. Umgekehrt findet man bei konstanter Kristall-

dicke t und einer Deformation des Gitters, die eine Änderung des Anregungsfehlers bedeutet, lokale Kontraständerungen. Das gemeinsame Merkmal einer Reihe solcher Defekte sind parallele helle und dunkle Streifen. Für den Dunkelfeldstrahlengang gelten analoge Überlegungen.
Eine systematische Einteilung der z. B. in Planardefekten entstehenden Streifentypen wurde von GEVERS, ART und AMELINCKX vorgenommen. Die Identifizierung von Kristalldefekten erfordert vielfach ein Modell, bei dem unter bestimmten Voraussetzungen der Abbildungsprozeß durch ein Differentialgleichungssystem computermäßig simuliert und mit der elektronenmikroskopischen Abbildung verglichen werden kann.
Das Auflösungsvermögen für Beugungskontrastuntersuchungen liegt bei ≈ 10 nm, und durch Arbeiten mit schwachen Strahlen (Dunkelfeld) bei ≈ 2 nm. Verständlich wird dieser Unterschied, weil bei der Dunkelfeldabbildung die absoluten Intensitäten meist geringer, die relativen Intensitätsunterschiede aber größer als bei der Hellfeldabbildung sind.

7.7.3. Elektronenmikroskopische Direktabbildung von Oberflächen mit dem Rasterelektronenmikroskop

Das gegenwärtig am häufigsten genutzte Gerät zur direkten Abbildung von Oberflächen mit elektronenoptischen Methoden ist das Rasterelektronenmikroskop. Abb. 7.7.3–1 zeigt den schematischen Aufbau eines Gerätes. Die Oberfläche des zu untersuchenden Objektes wird mit einer sehr feinen Elektronensonde, deren Durchmesser ≈ 10 nm beträgt, punktweise abgerastert. Die bei der Wechselwirkung zwischen Elektronensonde und dem jeweiligen Objektpunkt entstehenden Signale der verschiedensten Art werden von Detektoren erfaßt, über ein elektronisches Verstärkersystem verarbeitet und zur Hell-Dunkelsteuerung des synchron mit der Objektabrasterung auf einem Oszillografenschirm laufenden Elektronenstrahles benutzt. Eine Vergrößerung, die das Verhältnis des abgerasterten Bildbereiches zum abgerasterten Objektbereich kennzeichnet, kann durch geeignete Strahlablenkeinheiten in einfacher Weise elektronisch eingestellt werden.
Zur Informationsgewinnung über Objektbereiche können

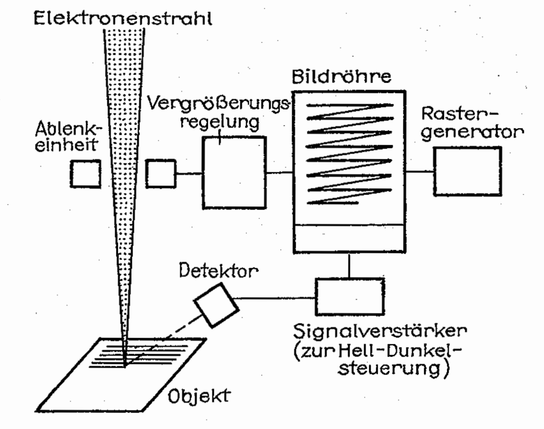


Abb. 7.7.3–1 Grundprinzip des Oberflächenrasterelektronenmikroskops

alle in der Abb. 7.7.1–1 aufgeführten Wechselwirkungen von Elektronen mit Materie genutzt werden. Für die Abbildung von Oberflächen, bei der das Interesse an einer hohen lateralen Auflösung im Vordergrund steht, werden wegen ihrer geringen Energie (< 50 eV) und der damit verbundenen kleinen Austrittstiefe vor allem die im Objekt ausgelösten Sekundärelektronen herangezogen. Die für Oberflächenuntersuchungen wichtigsten Kontrasteffekte sind der Topografie- bzw. Flächenneigungskontrast und der Materialkontrast, da Erhebungen oder geneigte Flächen Abschattungseffekte und geänderte Sekundäremissionseigenschaften hervorrufen. Außerdem unterscheiden sich die Elemente in ihren Sekundäremissionskoeffizienten. Die reflektierten Elektronen werden gleichfalls zur Erzielung von Material- und Topografiekontrasten ausgenutzt. Wegen ihrer wesentlich höheren Energie (10 bis 50 keV) und des damit verbundenen viel größeren Anregungsgebietes (vgl. Abb. 7.7.3–3) liegt allerdings die laterale Auflösung bei ≈ 100 nm. Eine Zusammenstellung einiger mit der Rasterelektronenmikroskopie gewinnbaren physikalischen Informationen sind mit ihrem Nachweissignal in Tab. 7.7.3–2 angegeben. Die Erfassung emittierter Röntgenstrahlung

Tab. 7.7.3–2 Physikalische Informationen aus rasterelektronenmikroskopischen Untersuchungen

Information	verwendetes Signal
Topografie und Geometrie einer Oberfläche	Sekundärelektronen (SE) rückgestreute Elektronen (RE)
kristalline Struktur (Channelling Diagramm)	SE, RE, absorbierte Elektronen
elektrische und magnetische Potentialverteilungen	SE, RE, Potentialkontrast
Rekombinationsprozesse in Halbleitern	Elektron-Loch-Paarerzeugung (Katodolumineszenz, EBIC)
Elementzusammensetzung (qualitativ und quantitativ)	Röntgen-Quanten (Mikroanalyse)

und ihre Analyse ist Gegenstand der analytischen Elektronenmikroskopie. Oberflächen-Rasterelektronenmikroskope haben für den Routinebetrieb eine Auflösung von 10 bis 20 nm. Unter günstigen Bedingungen sind Auflösungen bis ≈ 3 nm erzielbar. Dieser Wert ist weniger durch den Elektronensondendurchmesser begrenzt, da heute mit Feldemissionskatoden Sondendurchmesser von einigen Zehntel Nanometer möglich sind. Er hat seine Ursache vor allem darin, daß die zur Abbildung verwendeten Sekundärelektronen bei der hochauflösenden Rasterelektronenmikroskopie nicht nur aus dem unmittelbar von Elektronen direkt getroffenen Bereich stammen, sondern aus dem in Abb. 7.7.3–3 gekennzeichneten Hals des angrenzenden birnenförmig verbreiterten Gebietes, in das die Elektronen gestreut werden. Günstigere Bedingungen für die Auflösung liegen deshalb bei der Abbildung dünner Folien im Durchstrahlungsrastergebiet vor, der auch bei der Mikroanalyse kleinster Bereiche wesentliche Fortschritte gebracht hat.
Ein entscheidender Vorteil der Oberflächen-Rasterelektronenmikroskopie ist ihre große Schärfentiefe. Man kann mit

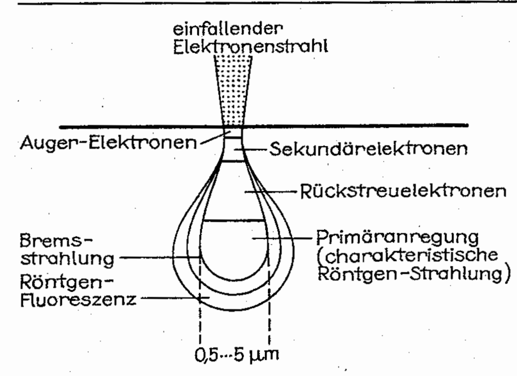


Abb. 7.7.3-3 Räumliche Verteilung der für den Nachweis verwendbaren Signale in einer massiven Probe

einem Wert von 100 µm bei etwa 1000facher Vergrößerung rechnen. Bereits diese Möglichkeit einer plastischen Abbildung sichert verständlicherweise die breite Anwendung dieser Methode, auch wenn sie die Lichtmikroskopie in der Auflösung nicht übertreffen würde. Sicherlich hat zur schnellen Verbreitung der Rasterelektronenmikroskopie der geringe Aufwand für die präparative Vorbereitung beigetragen, der besonders signifikant im Vergleich zur Durchstrahlungselektronenmikroskopie ist. Für nichtleitende Untersuchungsobjekte ist zur Vermeidung elektrischer Aufladungen lediglich eine dünne leitfähige Aufdampfschicht, z. B. aus Kohle, notwendig. Bei halbleitenden bzw. leitenden Untersuchungsobjekten kann meist ganz auf eine solche Präparation verzichtet werden.

7.7.4. Analytische Elektronenmikroskopie

Moderne Rasterelektronenmikroskope oder auch Durchstrahlungsgeräte werden häufig mit Röntgenspektrometern zusätzlich ausgerüstet, so daß neben den geschilderten Abbildungsverfahren zusätzlich eine Analyse der durch die Elektronen ausgelösten charakteristischen Röntgenstrahlen zur Information über die chemische Zusammensetzung der abgebildeten Gebiete möglich wird. Die Abb. 7.7.3-3 bringt eine schematische Darstellung der Anregungsbereiche einer massiven Probe, d. h. die räumlichen Quellen der zur Verarbeitung verfügbaren Signale. Die speziell für die analytische Elektronenmikroskopie maßgebenden Prozesse sind in der Abb. 7.7.4-1 in einem Energieniveauschema eines Atoms angegeben, das von Elektronen beschossen wird. In die weiteren Betrachtungen sollen nur die infolge der eingestrahlten Elektronen ausgelösten Röntgenstrahlen einbezogen und abschließend die primären Elektronen erwähnt werden, die einen Energieverlust infolge des Auslösens von Elektronen der inneren Schalen des bestrahlten Elementes erleiden. Für Oberflächenanalysen spielen die Auger-Elektronen eine wesentliche Rolle, sie liegen außerhalb dieser Betrachtungsweise.

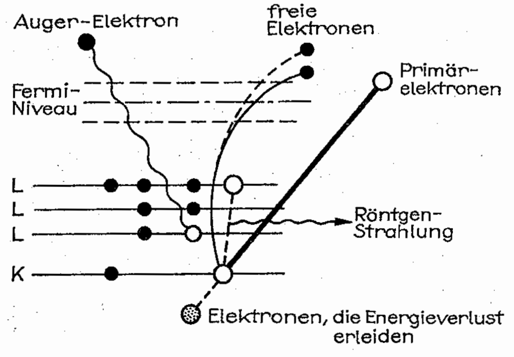


Abb. 7.7.4-1 Energieniveauschema eines Atoms mit Darstellung der bei Elektronenbeschuß stattfindenden Prozesse

7.7.4.1. Mikroanalyse von massiven Proben. Die Grundlage für die Ermittlung der qualitativen und quantitativen Elementzusammensetzung bildet die Analyse charakteristischer Röntgenstrahlen, die durch die Einwirkung energiereicher geladener Teilchen, speziell Elektronen, entstehen. Man unterscheidet für die spektrale Zerlegung der Röntgenstrahlen die wellenlängendispersive (WDS) und die energiedispersive Methode (EDS).

Der Vorteil des energiedispersiven Meßsystems besteht darin, daß das gesamte Energiespektrum simultan, sehr schnell (wenige Sekunden bis Minuten) und durch Kopplung mit elektronenoptischen Geräten von auswählbaren definierten Objektstellen aufgenommen werden kann. Der verwendete lithiumgetriftete Siliciumdetektor mit Berylliumfenster und einer Energieauflösung von 140 bis 150 eV bei 5,894 keV (Mangan K_α-Linie) begrenzt die Analyse auf Elemente ab der Ordnungszahl $Z > 11$ (Natrium). Unterhalb dieses Elementes muß dann auf das wellenlängendispersive System zurückgegriffen werden, das mit Hilfe verschiedener wechselbarer Kristalle den Nachweis von Elementen der Ordnungszahl $Z > 4$ gestattet. Die qualitative Analyse wird durch die angeschlossene Rechentechnik mit Speichereinheit außerordentlich erleichtert. Über einfache Befehle kann eine Elementidentifizierung vorgenommen werden, wobei die K-, L- und M-Linien in Abhängigkeit von der Wahl des Energiebereiches gleichzeitig mit der relativen Intensität auf einen Farbdisplay eingeblendet werden können. Der direkte Vergleich mit dem aufgenommenen Spektrum nach der Lage und Intensität ermöglicht eine eindeutige Elementidentifizierung.

Für die quantitative Analyse werden hauptsächlich drei Effekte berücksichtigt:

- die Ordnungszahlkorrektur (Z)
- der Absorptionseffekt (A)
- die Fluoreszenzanregung (F).

Das dafür verwendete Standardprogramm MAGIC V nach Colby kommt mit einem Minimum an Konstanten aus, die von einem Computersystem abgefragt werden. Nach Auswahl gespeicherter Standards wird die Korrektur berechnet

Tab. 7.7.4-2 Vergleich der Nachweisempfindlichkeiten von Elektronenstrahlmikroanalyse (WDS, EDS) und Elektronenenergieverlustanalyse (EELS)

Methode	Minimal nachweisbare Massenverhältnisse	Volumen	Masse
WDS	10^{-4}	1 (µm)³	10^{-15} g
EDS (SEM)	10^{-3}	1 (µm)³	10^{-14} g
(STEM)	$5 \cdot 10^{-3}$	10^4 (nm)³	10^{-17} g
EELS	10^{-2}	10^2 (nm)³	10^{-20} g

und mit dem Ausdruck der Ergebnisse automatisch abgeschlossen. Auch eine sogenannte standardlose Analyse (über auswählbares Programm) ist möglich, da für energiedispersive Systeme die Nachweiswahrscheinlichkeit eine einfache, leicht zu bestimmende Funktion der Energie ist. Diese Methode wurde von Russ entwickelt. Die „ohne" Standards ausgeführten Analysen sind vielfach von ausreichender Genauigkeit. Die Nachweisgrenzen für die Analyse von massiven Proben im Rasterelektronenmikroskop sind in Tab. 7.7.4–2 angegeben. Die laterale Auflösung ist infolge der birnenförmigen Anregung durch den Elektronenstrahl auf ≈ 1 µm begrenzt.

7.7.4.2. Mikroanalyse dünner Schichten. Zur Strukturbestimmung von Festkörpern nimmt das Transmissionselektronenmikroskop (TEM oder STEM) aufgrund seines lateralen Auflösungsvermögens, die heute eine Atomabbildung ermöglicht, eine Sonderstellung ein. Es existiert jedoch seit langem der Wunsch, neben geometrischen und kristallinen Feinstrukturen eine Elementanalyse durchzuführen. Für kristalline Untersuchungsobjekte kann die Elektronenbeugung als mikroanalytisches Verfahren gewertet werden, allerdings sind die Ergebnisse oft schwierig zu erhalten oder sogar mehrdeutig.

Einen entscheidenden Durchbruch brachte die Kombination des Durchstrahlungselektronenmikroskops mit einem energiedispersiven Röntgenspektrometer durch Cliff und Lorimer. Der Grundgedanke war, daß in durchstrahlbaren dünnen Schichten die primär erzeugten Röntgenstrahlen nur eine geringe Wahrscheinkeit haben, mit der Schicht selbst in Wechselwirkung zu treten. Damit braucht weder die Absorption noch die Fluoreszenz berücksichtigt werden. Für die Intensitätsverhältnisse I_1/I_2 zweier Proben ergibt sich der einfache Zusammenhang zu ihren Konzentrationsverhältnissen C_1/C_2 über $I_1/I_2 = k(\mathrm{E})\, C_1/C_2$, wobei die k elementspezifische Konstanten, sogenannte Cliff-Lorimer-Faktoren sind. Dafür existieren experimentell gemessene Werte, die besonders für schwere Elemente mit den berechneten gut übereinstimmen. Die Formel bezeichnet man als 1. Castaingsche Näherung. Empirische Korrekturen lassen hier weitere Verbesserungen erwarten. Die in STEM-Mode erzielbare räumliche Auflösung wurde von Doig, Lonsdale und Flewitt mit Annahme einer Normalverteilung für die Intensität als Funktion des Abstandes vom Strahlzentrum berechnet. Daraus können folgende Aussagen abgeleitet werden:

1. Höhere Spannungen liefern eine bessere räumliche Auflösung.
2. Die gemessenen Verteilungsprofile sind relativ unempfindlich bezüglich des Elektronensondendurchmessers, so daß dieser nicht zu klein gewählt werden braucht (≈ 10 nm wie der übliche Sondendurchmesser im Rasterelektronenmikroskop).

Die höhere Auflösung in STEM-Mode wird damit gegenüber der SEM-Mode sofort offensichtlich. Man rechnet mit einem Faktor 50 bis 100. Damit können Mikroanalysen aus durchstrahlbaren Proben mit einem mittleren Durchmesser von 10 bis 30 nm ausgeführt werden. Die relative Grenze für die Anzeige eines Elementes in STEM-Mode liegt bei 0,5 bis 1 Gewichtsprozent.

Eine Tiefenauflösung im üblichen Sinne existiert nicht. Die praktische Ausführung der Analyse unterscheidet sich wenig von der für massive Proben beschriebenen. Die Softwareentwicklung hat für TEM und STEM einen vergleichbaren Stand mit SEM erreicht. Für dünne biologische Objekte entwickelte Hall die quantitative Mikroanalyse.

Die genannte Gerätekombination TEM/STEM/SEM mit EDS war der Beginn der analytischen Elektronenmikroskopie. In den letzten Jahren hat sich die Elektronen-Energieverlustspektroskopie (EELS) als komplementäre Methode hinzugesellt, da hier die Anzeigeeffektivität für die leichtesten Elemente unterhalb Natrium am größten ist. Es können Flächen von ≈ 2 bis 3 nm ab Beryllium untersucht werden, so daß ein modernes analytisches Elektronenmikroskop alle Elemente mit der Ordnungszahl $Z \geqq 3$ nachzuweisen gestattet.

Als Beispiel sollen die mikroanalytischen Untersuchungen an einer Nickel/Gold-Schichtkombination auf einer Bronzeunterlage nach Bestrahlung mit einem Rubinlaser und nachfolgender Auslagerung der Probe in eine Schadgasatmosphäre aufgeführt werden. Ein aufgenommenes Energiespektrum der ausgelösten Röntgenstrahlung ist in Abb. 7.7.4–3 aufgeführt, dabei stellen die Peaks die charakteristischen Linien von im Untersuchungsgebiet vorhandenen Elementen dar, die über eine qualitative Analyse identifiziert wurden. Die mit Gold M-, Nickel K_α- und Kupfer K_α-Strahlung selektiv aufgenommenen Rasterbilder von der gleichen Probenfläche stellen die lokalen Verteilungen der jeweiligen Elemente dar. Das Beschußzentrum besteht vorwiegend aus Kupfer. Nickel kann dort nicht nachgewiesen werden, während am Rand und der Umgebung satellitenartig Zentren von Nickel sichtbar sind, die dem SE-Bild gut zugeordnet werden können. Für Gold kann ebenfalls die Verteilung festgestellt werden. Es ist beispielsweise in den Satelliten nicht existent. Die im Beschußzentrum schwach vorgetäuschte Verteilung stammt von Schwefel als Korrosionsprodukt ab, dessen K_α-Linie sich der Gold M-Linie überlagert. Hier kann durch Verwendung einer Gold L_α-Linie zur Abbildung eine Unterscheidung erfolgen.

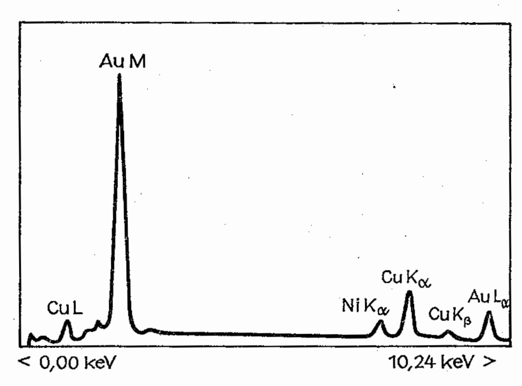


Abb. 7.7.4–3 Energieverteilung der emittierten Röntgen-Strahlung einer Ni/Au-Schichtkombination auf Bronze

7.8. Messung von und mit Röntgen- und Gammastrahlung

Röntgen- und Gammastrahlen (γ-Strahlen) sind *elektromagnetische Wellen.* Die Einordnung in das *elektromagnetische Spektrum* ist zwischen Wellenlängen von ≈ 5 nm bis ≈ 10^{-6} nm möglich.

Die Erzeugung von Röntgen- und γ-Strahlung ist mit energetischen Vorgängen im Atom bzw. dessen Kern verbunden. Bei der Charakterisierung der γ-Strahlung gibt man i. allg. nicht, wie bei Röntgenstrahlung üblich, die Wellen-

länge an, sondern ihre Energie. Die dabei noch vielfach verwendete Maßeinheit ist das Elektronenvolt [eV].
Drei Eigenschaften der Röntgen- und γ-Strahlung bilden im wesentlichen die Grundlage für die zerstörungsfreie Werkstoffprüfung und Strukturforschung: die *Absorption*, die Anregung von *Eigenstrahlung* und die *Beugung* in Kristallgittern. Bei der Aufklärung von biologischen und organischen Strukturen ist die Röntgenstrukturanalyse bis heute unübertroffen.

7.8.1. Nachweis von Röntgen- und γ-Strahlung

Der erste Nachweis beruhte auf der fluoreszierenden Wirkung der Röntgen-Strahlung (Röntgen 1895). Die erzeugte Röntgenstrahlung regte in der Nähe befindliche, fluoreszenzfähige Stoffe zum Leuchten an.
γ-Strahlung wurde erstmals von Becquerel (1896) auf einer Fotoplatte nachgewiesen (wie auch schon kurze Zeit vorher von Röntgen die Röntgen-Strahlung). Das war möglich, weil beide Strahlungen die Eigenschaft haben, fotografische Platten (Filme) zu „belichten". Diese Methode hat sich für beide Strahlungen als Nachweis bis heute erhalten (Kernspurplatten, Röntgen-Filme). Alle weiteren Nachweismethoden oder -geräte (Detektoren) beruhen auf der ionisierenden Wirkung der Strahlungen.
Geiger-Müller-Zählrohr (Auslösezählrohr). Es ist der bekannteste Detektor. Die Röntgen- oder γ-Strahlung dringt durch ein möglichst schwach absorbierendes Fenster in ein gasgefülltes Metallrohr ein und ionisiert das darin befindliche Füllgas (Argon, Xenon, Methan). Die erzeugten Ionen führen zu einer Stromleitung zwischen der Rohrwandung und einem in der Mitte angebrachten Draht. Am Widerstand R fällt eine Spannung ab, die elektronisch weiter verarbeitet wird (Abb. 7.8.1–1). Die Größe der erzeugten Spannungsimpulse ist von der Energie der einfallenden Strahlung unabhängig und kann mit einem elektronischen Meßgerät registriert werden.
Eng mit dem Geiger-Müller-Zählrohr verwandt ist das *Proportionalzählrohr.* Auch der Aufbau ist im wesentlichen der gleiche. Der Unterschied besteht darin, daß die vom Detektor abgegebene Impulshöhe von der Energie der einfallenden Strahlung abhängig ist. Man kann die registrierte Strahlung nach ihrer Energie „sortieren".
Szintillationszähler. Er gilt neben dem Nachweis von Röntgen- und γ-Strahlung mit Hilfe einer fotografischen Platte als einer der ältesten Detektoren. Die Strahlung fällt auf einen fluoreszenzfähigen Kristall (z. B. ZnS). Die entstehenden Lichtblitze oder Szintillationen kann man zählen oder anderweitig registrieren, wobei die Folge der zu registrierenden Quanten nicht zu dicht sein darf.

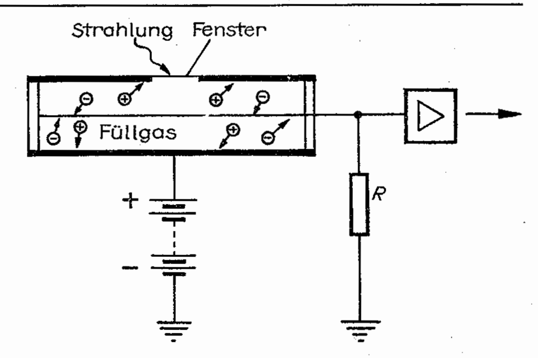


Abb. 7.8.1–1 Geiger-Müller-Zählrohr

Sekundärelektronenvervielfacher (SEV) haben diesen Nachteil nicht. Die zu messende Röntgen- oder γ-Strahlung fällt auf den fluoreszenzfähigen Kristall (*Szintillator*, z. B. $CaWO_4$, NaJ). Die entstehenden Lichtblitze gelangen über ein optisches System auf die Fotokatode des SEV. Diese wandelt die Lichtenergie in kinetische Energie in Form von Fotoelektronen um (*Fotoeffekt*). Die Fotoelektronen werden durch eine angelegte Spannung beschleunigt. Dadurch erhalten sie eine höhere Energie. Sie sind nun in der Lage, eine größere Anzahl von Elektronen (Sekundärelektronen) beim Auftreffen auf die erste Prallelektrode (*Dynode*) des SEV herauszulösen. Dieser Vorgang wird einige Male wiederholt und man gelangt zum Schluß zu einer sehr großen Anzahl von Elektronen am Ausgang des SEV (Abb. 7.8.1–2). Über einen nachgeschalteten Verstärker ist die Registrierung des Ausgangssignals möglich.
Halbleiterdetektoren werden wegen ihres hohen Energieauflösungsvermögens immer häufiger eingesetzt. Den Aufbau zeigt Abb. 7.8.1–3. Zwischen einer p- und n-leitenden Siliciumschicht befindet sich eine mit Lithium dotierte eigenleitende Schicht (intrinsic oder i-Gebiet). Das in das i-Gebiet eindringende Röntgen- oder Gammaquant erzeugt über den Fotoeffekt freie Ladungsträger. Dadurch kommt es zu einem Stromfluß durch den Detektor. Die Dauer des Stromstoßes liegt bei einigen Nanosekunden. Nachteilig ist, daß der Detektor zur Zeit oft noch mit flüssigem Stickstoff ($T \approx 77$ K) gekühlt werden muß. Bei Experimenten mit Röntgen-Strahlen, bei denen man Informationen über die Verteilung der gebeugten Strahlung um das Untersuchungsobjekt (*Diffraktometrie*) gewinnen möchte, werden verstärkt ringförmig gebaute Zähler verwendet. Damit entfällt die mitunter komplizierte Mitbewegung des Detektors. Der Aufbau ist so vorstellbar, als ob der Gesamtring aus lauter kleinen Halbleiterdetektoren zusammengesetzt wäre. Oft

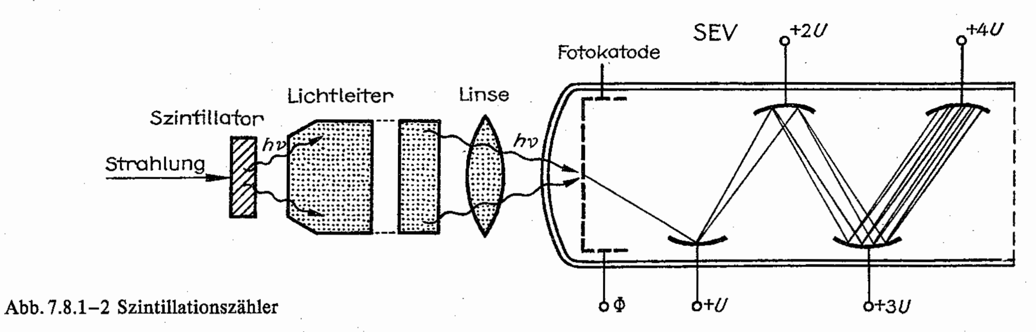


Abb. 7.8.1–2 Szintillationszähler

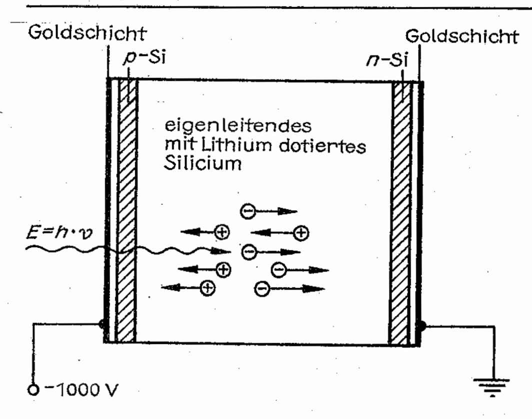


Abb. 7.8.1–3 Halbleiterdetektor

sind diese kleinen Zählsegmente aus noch kleineren Einheiten zusammengesetzt, deren jede nur für einen bestimmten Wellenlängenbereich (energie-)empfindlich ist. Dadurch ist es möglich, sowohl den Ort als auch den Energiebereich der Strahlung gleichzeitig zu registrieren. Bei einigen Experimenten sind nur der Ort und die Intensität der Strahlung wichtig, so beim *Laue-Verfahren.* Für diesen Zweck wurde ein „Multiwire"-Detektor (Proportionalzählrohr mit mehreren Zähldrähten) entwickelt, der die sofortige Darstellung eines *Laue-Diagramms* auf einem Fernsehmonitor erlaubt. Dieser flächenhaft aufgebaute Detektor ersetzt den sonst üblichen Film. Die Grundidee besteht darin, daß der Detektor über einen Verstärker mit einem Monitor so verbunden ist, daß die wirksame Zählfläche des Detektors der Fläche des Bildschirms entspricht. Durch eine geeignete Anordnung der Detektorzähldrähte untereinander besteht die Möglichkeit, die unterschiedliche Intensität der Röntgen-Strahlung, bei der Einwirkung auf den Detektor, sofort auf den Monitor als Bildkontrast abzubilden.

Nebel- und Blasenkammern benutzt man vorzugsweise zum Nachweis von Kernstrahlung (vgl. 5.4.4.4.).

7.8.2. Messen von Kenngrößen der Röntgen- und γ-Strahlung

7.8.2.1. Messen der Wellenlänge. Erst etwa 30 Jahre nach der Entdeckung der Röntgen-Strahlung war man in der Lage, die Wellenlänge mit hinreichender Genauigkeit zu messen. Dies erfolgte an einem *Beugungsgitter* mit streifender *Inzidenz* (Abb. 7.8.2–1) $\lambda z = (g/2) \cdot \Delta \cdot \gamma$ (z = Ordnung des Beugungsmaximums). Weitere Wellenlängenmessungen sind durch Röntgen-Beugungsexperimente an Einkristallen möglich. Dazu ist die Kenntnis der *Gitterkonstante* eines Untersuchungskristalls mit entsprechender Genauigkeit notwendig. Diese kann aus der Kristallgeometrie, dem *Molekulargewicht,* der *Dichte* und der *Avogadro-Zahl* ermittelt werden. Eine weitere Methode besteht in der relativen Messung der gesuchten Wellenlänge zu einer anderen geeigneten Wellenlänge (*Emissions-* oder *Absorptionslinie*), deren Wert bereits genau genug bekannt ist.

Zur Energiebestimmung von γ-Strahlung gibt es ebenfalls verschiedene Methoden. Für Energien von *γ-Quanten,* die 1 MeV ($\lambda = 0{,}001$ nm) nicht wesentlich überschreiten, benutzt man i. allg. die von Du Mond entwickelte Methode der *Kristallgitterspektroskopie.* Bei Kenntnis der *Gitterkonstante* des Meßkristalls und nach Messung des Beugungswinkels der γ-Strahlung, kann über die *Braggsche Beziehung* $2d \cdot \sin \Theta = n \cdot \lambda$ (d = Netzebenenabstand, n = Beugungsordnung) die Wellenlänge bestimmt werden. Bei allen anderen Methoden (vor allem bei Energien größer als 1 MeV) nutzt man die Möglichkeit der Energieübertragung von den γ-Quanten auf Elektronen aus. Die Elektronenenergie kann man mit *magnetischen Spektrografen* bestimmen. Die Energieumwandlung kann beispielsweise über den *Compton-* oder den *Fotoeffekt* erfolgen.

7.8.2.2. Messen der Intensität. Unter der Intensität einer Strahlung versteht man das Verhältnis der auf eine Fläche auftretenden *Strahlungsenergie W* zur Größe dieser *Fläche A* und der *Bestrahlungszeit t,* $I = W/(A \cdot t) = P/A$. Zum Beispiel mißt man unter einem Raumwinkel von 2π und in einer Entfernung von 100 mm von einer mit 1 bis 2 kW elektrischer Leistung betriebenen Röntgenröhre eine Intensität von $I \approx 16\,\mathrm{Wm}^{-2}$. Um die Energie der Strahlung zu bestimmen, ermittelt man die *integrale Intensität,* die der Fläche einer registrierten Meßkurve proportional ist. Für Vergleichszwecke mißt man oft nur *relative Intensitäten* in einer beliebig gewählten Einheit. Nur bei speziellen Problemen benötigt man *absolute Intensitäten,* d. h. den Bruchteil der Primärstrahlintensität, der vom Untersuchungskristall bei einem bestimmten *Braggwinkel* Θ gestreut wird.

7.8.2.3. Andere Kenn- und Meßgrößen. Auch bei Röntgen- und γ-Strahlung tritt *Dispersion* auf, d. h. die Abhängigkeit der *Brechzahl n* von der Wellenlänge der Strahlung. Die Brechzahl n weicht sehr wenig von 1 ab, $\delta = (1 - n) \approx 1{,}2 \cdot 10^{-5}$. Der Brechungsindex δ hängt von der Dichte ϱ des durchstrahlten Materials, der verwendeten Wellenlänge λ und dem Verhältnis der Summe der Ordnungszahlen zu der Summe der relativen Atomgewichte aller Atome der Elementarzelle des Untersuchungsmaterials ab, $\delta = 27{,}1\,\mathrm{mkg}^{-1} \cdot \lambda^2 \cdot \varrho \cdot (\sum Z / \sum A)$.

Eine weitere sehr wichtige Größe ist die Absorption. Durchdringt Röntgen- oder γ-Strahlung der Wellenlänge λ ein Material der Dicke d, dann wird die einfallende Strahlung der Intensität I_0 um den Faktor $e^{-\mu d}$ geschwächt, so daß man nur noch die Intensität $I = I_0 e^{-\mu d}$ registriert. Der *Schwächungskoeffizient* μ setzt sich aus drei Anteilen zusammen: dem *Absorptionskoeffizient* τ, dem *Streuungskoeffizient* σ und dem *Paarbildungsbeiwert* $\varkappa$, $\mu = \tau + \sigma + \varkappa$. In der Spezialliteratur ist i. allg. der *Massenschwächungskoeffizient* μ/ϱ für jedes Element bei einer bestimmten Wellenlänge λ tabelliert (ϱ = Dichte des Materials). Bei Gemengen aus verschiedenen chemischen Elementen berechnet sich der Massenschwächungskoeffizient nach $(\mu/\varrho)_{\mathrm{Gemenge}} = \sum_i \alpha_i (\mu_i/\varrho_i)$ (α_i = Gewichtsanteil der Komponente i).

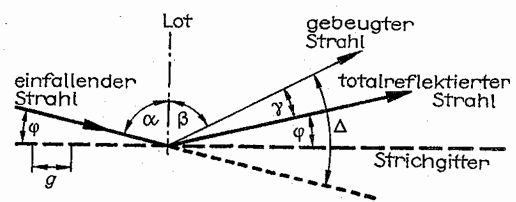


Abb. 7.8.2–1 Anordnung zur absoluten Bestimmung der Wellenlänge von Röntgen-Strahlung (Gitterkonstante $g = 5 \cdot 10^{-4}$ cm)

Die Eigenschaft der Röntgen-Strahlung, von verschiedenen Stoffen unterschiedlich stark geschwächt zu werden, wird bei allen Verfahren der medizinischen *Röntgen-Diagnostik* ausgenutzt, z. B. Schirmbildaufnahme, Schichtaufnahmetechnik (Tomografie), Computertomografie.

7.8.3. Methoden zu Werkstoff- und Kristalluntersuchungen unter Verwendung von Röntgen- und γ-Strahlung

7.8.3.1. Grobstrukturuntersuchungen. Sowohl in der Forschung als auch in der Industrie ist es notwendig, die für wissenschaftliche Zwecke oder technische Prüfungen zur Verfügung stehenden Proben oder Werkstücke mit *zerstörungsfreien Verfahren* zu untersuchen.

Röntgen-Grobstrukturverfahren. Röntgen-Strahlen durchdringen alle Stoffe mit mehr oder weniger großer Abschwächung der Intensität (vgl. 7.8.2.3.). Die Durchdringung ist abhängig von der Dicke des Werkstückes, seiner chemischen Zusammensetzung und der zur Untersuchung verwendeten Röntgen-Strahlung. Je dünner das Probenmaterial, je niedriger die Ordnungszahl des Elementes aus dem es besteht und je kleiner die Wellenlänge der Röntgen-Strahlung, um so besser ist die Durchdringung. So kann man mit Röntgen-Strahlung der Wellenlänge $\lambda = 0{,}0062$ nm mit gutem Erfolg eine Stahlplatte der Dicke $d = 60$ mm untersuchen. Angewendet wird die Grobstrukturuntersuchung hauptsächlich in der Industrie zur Materialprüfung, z. B. bei der Kontrolle von Schweißnähten auf ihre Güte oder zur Qualitätskontrolle von gegossenen und gezogenen Werkstücken. Das zu untersuchende Objekt wird von Röntgen- oder γ-Strahlung getroffen und der durch die Probe hindurchgedrungene Anteil schwärzt einen Film. Enthält das Prüfungswerkstück Materialfehler, z. B. einen Schlackeeinschluß, so ist für diese Störung der *Schwächungskoeffizient* ein anderer als für das umliegende Material. Auf dem Film führt das dazu, daß sich der Fehler entweder durch eine stärkere (kleinerer Koeffizient μ) oder geringere (größerer Koeffizient μ) Schwärzung gegenüber der Umgebung abhebt. Beim Einsatz von γ-Strahlern ist infolge der größeren Energie auch das Durchstrahlen dickerer Materialien möglich. Nachteilig wirkt sich die erheblich geringere Intensität der γ-Strahler gegenüber Röntgen-Strahlungsquellen aus, was bis zu 100fach längeren Belichtungszeiten führt. Grobstrukturanlagen sind auch ambulant einsetzbar.

7.8.3.2. Feinstrukturuntersuchungen. Röntgen-Feinstrukturuntersuchungen dienen zur Aufklärung des mikroskopischen Aufbaus (Strukur) eines bestimmten Materials. Trifft ein Röntgen-Strahl das zu untersuchende Material, so treten Beugungserscheinungen auf (vgl. 2.5.2.4.). In Abhängigkeit von der geometrischen Anordnung der Atome (Abstand, Regelmäßigkeit, Symmetrie), ihres Streuvermögens und der Wellenlänge der Strahlung beobachtet man unter verschiedenen Betrachtungswinkeln Intensitätsmaxima. Aus diesen Beugungsbildern bzw. -informationen lassen sich die gewünschten Rückschlüsse auf die atomaren Struktureigenschaften ziehen. Man kann die Feinstrukturuntersuchungen auf zwei verschiedene Arten klassifizieren. Nach den verwendeten Untersuchungsmaterialien unterscheidet man zwischen der Beugung *(Diffraktion)* am *Einkristall*, am *polykristallinen Material* und an *amorphen* Substanzen, Flüssigkeiten und *Flüssigkristallen*. Andererseits ist eine Einteilung nach geforderten oder gewünschten Strukturparametern möglich. Hiernach unterscheidet man zwischen *Ideal- (Strukturanalyse)* und *Realstrukturuntersuchungen*.

Laue-Verfahren. Die aus dem *Bremsspektrum* (einschließlich *charakteristischer Strahlung*) bestehende Röntgen-Strahlung fällt auf den zu untersuchenden Einkristall. Auf einem Röntgen-Film, der vor *(Rückstrahl-Verfahren)* oder hinter dem Kristall *(Durchstrahl-Verfahren)* stehen kann, wird das Beugungsdiagramm abgebildet. Dieses sogenannte *Laue-Diagramm* zeigt die gleiche Symmetrie wie der mikroskopische Kristallaufbau.

Drehkristall-Verfahren. Hierbei wird unter Ausnutzung *charakteristischer Strahlung* ein Einkristall um eine feste Achse gedreht. Die gebeugte Strahlung fällt unter dem Bragg-Winkel Θ auf einen den Kristall zylindrisch umgebenden Film. Diese Methode wird i. allg. zur Bestimmung der Gitterparameter genutzt. Dazu werden die auf dem Film erzeugten Reflexe mit den *Millerschen Indizes* (hkl) versehen *(Indizierung)* und über die *Braggsche Gleichung* läßt sich die Gitterkonstante z. B. eines *kubischen Gitters* bestimmen, $a = (n \cdot \lambda) \cdot (h^2 + k^2 + l^2)^{1/2}/(2 \cdot \sin \Theta)$. Außerdem gewinnt man wichtige kristallografische und strukturelle Informationen (Raumgruppe, Auslöschungen, Kristallzustand). Vom Drehkristallverfahren gibt es mehrere Abwandlungen, z. B. das *Schwenkverfahren*, das *Präzessionsverfahren* und das *Weissenberg-Verfahren*.

Debye-Scherrer-Verfahren. Bei ihm wird der Einkristall des Drehkristallverfahrens durch *polykristallines Material* ersetzt. Dieses Verfahren dient fast ausschließlich zur Bestimmung von Gitterparametern. Statt „Reflexpunkte" wie beim Drehkristallverfahren, erhält man hier ein Linienspektrum. Mit dem gleichen Verfahren lassen sich Flüssigkeiten, *Flüssigkristalle* und *amorphe Festkörper* untersuchen. Bei den erwähnten Methoden werden in zunehmendem Maß statt des Films die bereits erwähnten *elektronischen Zähler* (vgl. 7.8.1.) eingesetzt. Der Detektor läßt sich über elektronische Steuerungen an jeder beliebigen Stelle positionieren. In diesen Fällen spricht man von *Röntgen-* und *γ-Diffraktometern*, die heute vielfach mit Kleinrechnern gekoppelt sind. Das hat den Vorteil einer schnellen Datenauf- und -verarbeitung.

Kristall- oder Idealstrukturanalyse. Ihre Aufgabe ist die Bestimmung der Koordinaten der einzelnen Atomlagen $(x_r\, y_r\, z_r)$ in der *Elementarzelle*. Grundlage dafür ist die Berechnung der *Elektronendichteverteilung* $\varrho(xyz)$, wobei die Schwerpunkte der Atome als Maximum ausgewiesen sind.

Wegen der periodischen Anordnung der Atome im Kristall ist die Elektronendichteverteilung eine periodische Funktion, die sich durch eine Fourier-Synthese darstellen läßt:

$$\varrho(xyz) = 1/V \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} F(\text{hkl})\, e^{-2\pi i (hx + ky + lz)}$$

Dabei ist V das Volumen der Elementarzelle und $|F_{hkl}|$ die *Strukturamplitude*, die proportional der Wurzel aus der integralen Intensität des Reflexes mit der *Millerschen Indizes* (hkl) ist. F_{hkl} ist der *Strukturfaktor*, der mit den Atomkoordinaten durch die Beziehung $F_{hkl} = \sum_{r=1}^{N} f_r \cdot e^{2\pi i (hx_r + ky_r + lz_r)}$ verknüpft ist (f_r = Streufähigkeit eines Atoms).

Experimentell erhält man nur die Beträge $|F_{hkl}|$ aus der integralen Intensität der Reflexe. Die Strukturamplituden zweier Reflexe können also z. B. gleich sein (gleiche Intensität), während sich die Strukturfaktoren F_{hkl} durch die Vor-

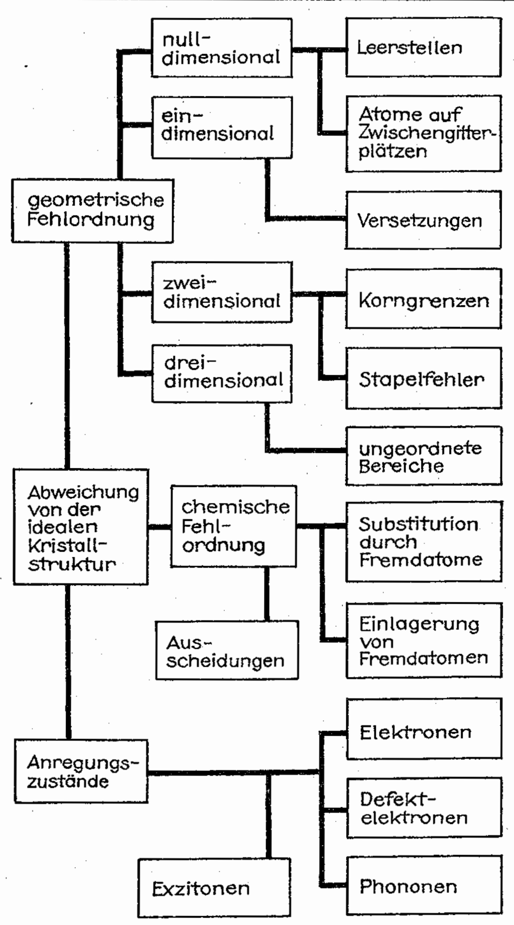


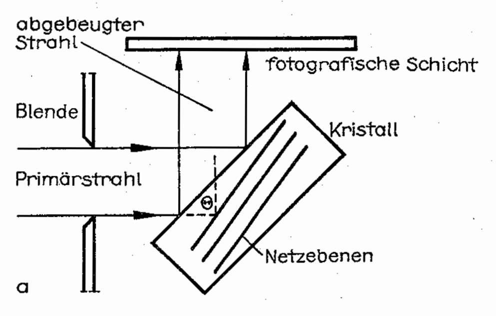
Abb. 7.8.3–1 Auswahl der wichtigsten Störungen der idealen Kristallstruktur

zeichen und damit durch ihre Koordinaten $(x_r y_r z_r)$ unterscheiden. Die Ermittlung des Vorzeichens mit Hilfe des Experiments ist sehr schwierig. Ursache dafür sind die unbekannten Phasenbeziehungen bei der Röntgenbeugung. Hier könnte z. B. der Einsatz eines Röntgen-Lasers sinnvoll sein.

Die Annahme der Existenz eines Idealkristalls bei der *Strukturanalyse* steht nicht damit im Widerspruch, daß es einen solchen Kristall nicht gibt (Idealkristall: dreidimensionale, bis ins Unendliche fortgesetzte periodische Anordnung von Atomen oder Atomgruppen). Jede Abweichung vom Konzept des Idealkristalls führt zum *Realkristall*, mit dessen Untersuchung sich die *Realstrukturmethoden* befassen. Diese Untersuchungen sind deshalb so wichtig, da sich unter dem Einfluß schon geringer Strukturstörungen viele Kristalleigenschaften beträchtlich ändern können, wie die Lumineszenz, die Festigkeit, die Leitfähigkeit u.a. Eine solche Störung liegt z. B. bei Halbleiterkristallen vor, wenn sie mit Fremdatomen dotiert werden, um bestimmte elektronische Eigenschaften zu erzielen. Einige der wichtigsten Kristallstörungen sind in Abb. 7.8.3–1 dargestellt.

Röntgen-Topografie. Unter gewissen Voraussetzungen eignen sich die bisher erwähnten Methoden auch zu Realstrukturuntersuchungen. Allerdings liefert die *Röntgen-Diffraktometrie* nur indirekte oder integrale Aussagen über die Kristalldefekte. Es erhebt sich daher die Frage, ob es auch möglich ist, genügend große *Gitterfehler* (Strukturstörungen) auch direkt abzubilden. Für die Untersuchung von z. B. *Versetzungen, Stapelfehlern* und *Korngrenzen* (vgl. Abb. 7.8.3–1) eignen sich die Methoden der Röntgen-Topografie. Darunter versteht man die Abbildung eines Einkristalls (oder eines Teils davon) mit Hilfe von an einer *Netzebenen*schar des Kristalls abgebeugter Röntgen-Strahlung. Dabei kann die Strahlung an dem betreffendem Kristall „reflektiert“ *(Reflextopografie)* oder er kann „durchstrahlt“ *(Transmissionstopografie)* werden (Abb. 7.8.3–2). Die Kristalldefekte bilden sich auf einem Film, dem sogenannten *Röntgen-Topogramm* durch unterschiedliche Schwärzungskontraste ab. Im Prinzip „fotografiert“ die Röntgen-Topografie den Kristall mit Hilfe der Röntgen-Strahlung und läßt auf dem Film die Kristallfehler erkennen. Für Feinstrukturuntersuchungen wird in zunehmendem Maß *Synchrotronstrahlung*, die z. B. beim Betrieb von *Elektron-Positron-Speicherringen* entsteht, eingesetzt. Mit Synchrotronstrahlungsquellen erreicht man zur Zeit eine etwa 1000fache höhere Röntgen-Intensität gegenüber herkömmlichen Quellen. Damit verringern sich die Zeiten zur Registrierung eines *Laue-Diagramms* oder einer *Drehkristallaufnahme* bis auf Bruchteile von Sekunden.

7.8.3.3. Spektroskopische Untersuchungen. Man unterscheidet zwischen qualitativer und quantitativer Spektroskopie. Die Aufgabe der *qualitativen Spektroskopie* ist z. B. der Nachweis eines bestimmten Elementes und die der *quantitativen Spektroskopie* die mengenmäßige Bestimmung eines Elementes in einer Probe. Untersucht werden sowohl



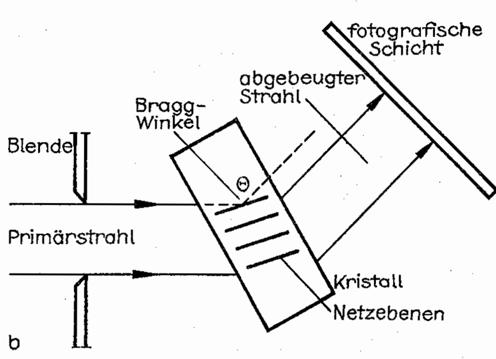


Abb. 7.8.3–2 Röntgentopografische Methoden, (a) Reflex- und (b) Transmissionstopografie

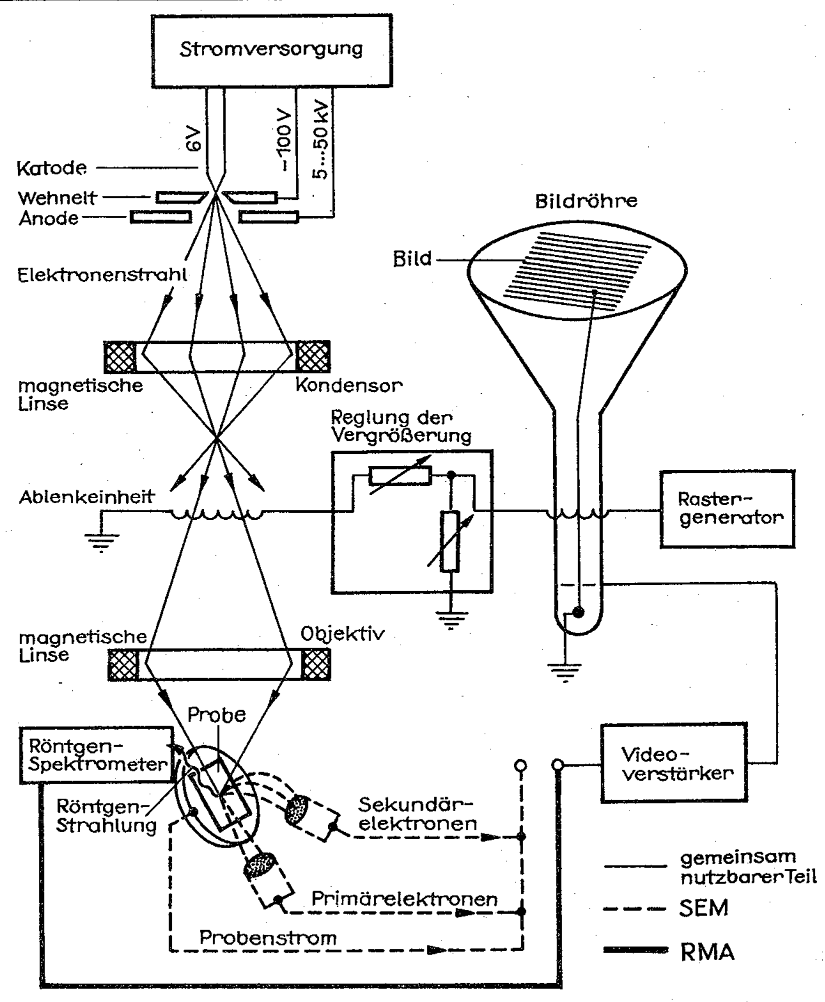


Abb. 7.8.3–3 Prinzipieller Aufbau und Wirkungsweise des Rasterelektronenmikroskops und des Röntgen-Mikroanalysators

Emissions- als auch *Absorptionsspektren.* Röntgen- oder γ-spektroskopische Untersuchungen können weiterhin Aussagen über Bindungsarten, -energien und -winkel liefern.

Die spektroskopische Untersuchung eines Emissionsspektrums setzt voraus, daß das Untersuchungsobjekt selbst das Spektrum aussendet. Dies ist aber nur in wenigen Fällen gesichert. Nur die *radioaktiven Elemente* senden ohne äußere Anregung γ-Strahlung aus. Natürliche Röntgen-Strahlungsquellen sind auf der Erde unbekannt. Demzufolge ist es bei der Untersuchung des Emissionsspektrums erst einmal notwendig, eine entsprechende Röntgen- oder/und γ-Strahlung anzuregen. Dies ist auf verschiedene Art und Weise möglich.

Elektronenstrahlmikroanalyse. Bei ihr werden Elektronen bestimmter Energie auf die Probe geschossen (Abb. 7.8.3–3). Diese Energie regt die *charakteristische Röntgen-Strahlung* der in der Probe enthaltenen Elemente an. Beim *wellenlängendispersiven Spektrometer* fällt die Röntgen-Strahlung auf einen *Monochromatorkristall* und bei bestimmten Kristallstellungen wird nach dem *Braggschen Gesetz* jeweils nur eine Wellenlänge reflektiert. Der Winkel wird gemessen und mit den bekannten Daten des Monochromatorkristalls die Wellenlänge berechnet. Aus dieser läßt sich dann auf das chemische Element schließen. Beim *energiedispersiven Spektrometer* wird die in den Detektor einfallende Röntgen-Strahlung wieder in Elektronen-Loch-Paare umgewandelt. Dadurch können Impulse verschiedener Impulshöhe, -form und -breite erzeugt werden, die einer bestimmten, charakteristischen Energie des Röntgen-Spektrums eines bestimmten Elementes zugeordnet werden können. Wird die Röntgen-Strahlung durch Beschuß von Ionen erzeugt, spricht man von *Ionenstrahlmikroanalyse.*

Zwischen dem *Röntgen-Mikroanalysator* (RMA) und den *Elektronenmikroskopen*, speziell dem *Rasterelektronenmikroskop* (SEM, engl. Scanning-Electron-Microscop), besteht eine große Ähnlichkeit. Deshalb ist es auch üblich, beide Methoden in einem Gerät zu vereinigen. Die prinzipielle Wirkungsweise beider Methoden ist in Abb. 7.8.3–3 dargestellt. Beim RMA wird die erzeugte Röntgen-Strahlung über einen Monochromatorkristall dem Detektor zur quali-

tativen und quantitativen Analyse zugeführt. Bei einem Elektronenstrahldurchmesser von 50 nm bis 1 µm und einem Energieauflösungsvermögen des Monochromatorkristalls für Röntgen-Strahlung von 20 eV sind Elementanalysen für Ordnungszahlen $Z \geqq 3$ möglich.

Röntgen-Fluoreszenzanalyse. Bei der *Spektroskopie der Röntgen-Fluoreszenz* wird die zu untersuchende Probe mit Röntgen-Strahlung hoher Energie (meist Wolframanode: $\lambda = 0{,}02171$ nm) bestrahlt. Dadurch werden die in der Probe enthaltenen Elemente zur Röntgen-Eigenstrahlung, der *Fluoreszenz* angeregt. Dies ist nur dann möglich, wenn die anregende Röntgen-Strahlung kurzwelliger (energiereicher) oder mindestens von der gleichen Wellenlänge ist, wie eine mögliche *Röntgen-Emissionslinie* des Elementes in der Probe. Die Spektroskopie der Fluoreszenzstrahlung erfolgt entweder mit einem wellenlängen- oder/und energiedispersiven Spektrometer.

Andererseits ist es auch möglich, durch Röntgen-Strahlung entsprechender Energie Elektronen aus dem Atomverband der in der zu untersuchenden Probe enthaltenen Elemente herauszulösen. Bei der „*E*lectron *S*pectroscopy for *C*hemical *A*nalysis" (ESCA) werden die auftreffenden Photonen in der Probe absorbiert. Jedes absorbierte Photon überträgt seine Energie vollständig auf ein Elektron, welches eine charakteristische Bindungsenergie im Festkörper aufweist. Ist die Photonenenergie größer als die Bindungsenergie, so verläßt das Elektron die Probe mit einer bestimmten kinetischen Energie. Die Registrierung der Elektronen erfolgt entsprechend ihrer Energie. Nach der Energie der Elektronen schließt man auf chemische und physikalische Eigenschaften der Probensubstanz.

Ein Beispiel für die Spektroskopie von γ-Strahlung ist die *Mössbauer-Spektroskopie* (*Mössbauer-Effekt*, vgl. 5.2.4.4.).

7.8.4. Strahlenschutz

Bei der Absorption von Röntgen- oder γ-Strahlung entstehen primär energiereiche Fotoelektronen, welche durch Stoßionisation eine große Zahl von Sekundärionen erzeugen. Erfolgt die Absorption in biologischen Geweben, so führt das zu Hautverbrennungen, Blutzersetzung oder Schädigung der inneren Organe. Deshalb ist jeder, der mit ionisierender Strahlung arbeitet, verpflichtet, sich entsprechend den Strahlenschutzbestimmungen zu schützen. Folgende Größen sind für den Strahlenschutz wichtig:

Aktivität von Radionukliden. Darunter versteht man die Anzahl der Zerfälle je Zeiteinheit. Die SI-Einheit für die Aktivität ist das Becquerel [Bq].

Expositionsdosis. Sie gibt die von einer ionisierenden Strahlung erzeugte Ladung in einem bestimmten Material an. Die SI-Einheit der Expositionsdosis ist das Coulomb je kg [C/kg]. Früher war dafür das Röntgen gebräuchlich, $1\,\text{R} = 2{,}58 \cdot 10^{-4}\,\text{C/kg}$.

Wenn die Expositionsdosis je Sekunde um 1 C/kg anwächst, so liegt die *Expositionsdosisleistung* von 1 A/kg vor.

Energiedosis. Sie kennzeichnet die Energie, die einem Körper durch ionisierende Strahlung zugeführt wird. Die SI-Einheit ist das Gray [Gy]. Früher war das Rad [rd] gebräuchlich, $1\,\text{rd} = 10^{-2}\,\text{Gy}$.

Äquivalentdosis. Für den Strahlenschutz ist die relative *biologische Wirksamkeit* (RBW) wichtig. Diese kann durch die *Äquivalentdosis* ausgedrückt werden. Die SI-Einheit ist das [Gem] (*G*ray *e*quivalent *m*an). Daneben ist die SI-fremde Einheit [rem] (*R*öntgen *e*quivalent *m*an) zulässig, $1\,\text{Gem} = 10^2\,\text{rem}$ (für Röntgen- und γ-Strahlung).

Strahlenexponierte Menschen unterliegen einer ständigen Kontrolle und medizinischer Betreuung.

8. Angewandte Elektrizitätslehre

8.1. Diskrete elektrische und magnetische Bauelemente

8.1.1. Widerstände, Kondensatoren, Spulen

Widerstände, Kondensatoren und Spulen sind passive Bauelemente. Das sind Zweipole, die elektrische Energie innerhalb eines endlichen Zeitintervalls entweder nur verbrauchen, d. h. in Wärme umwandeln oder nur speichern können. Für sie ist in jedem Punkt der Kennlinie das Produkt aus dem Momentanwert der Spannung und dem Momentanwert des Stromes größer oder gleich null. Im Gegensatz zu aktiven Zweipolen können passive Zweipole keine Hilfsenergie so in elektrische Energie umsetzen, daß die an den Anschlußklemmen entnehmbare elektrische Energie größer ist als die zugeführte.

8.1.1.1. Widerstände. Als elektrischen Widerstand bezeichnet man sowohl eine physikalische Größe wie auch ein Bauelement, das durch einen bestimmten Wert dieser Größe bestimmt ist. Die Dimensionierung der Widerstandswerte, die früher nach einer Auswahl aus den natürlichen Zahlen erfolgte, richtet sich gemäß einer internationalen Normung nach geometrischen Reihen. Die Quotienten der verwendeten Reihen sind in der Reihenfolge sinkender zulässiger Toleranzen (±20 %, ±10 %, ±5 %, ±2 %, ±1 %) $\sqrt[6]{10}$, $\sqrt[12]{10}$, $\sqrt[24]{10}$, $\sqrt[48]{10}$, $\sqrt[96]{10}$. Damit ist der Quotient benachbarter Widerstandswerte einer solchen Reihe eine Konstante.

Nach den Einsatzanforderungen der Widerstandsbauelemente, die sich hinsichtlich Alterungsbeständigkeit, Temperatureinfluß, Eigenschaften bei hohen Frequenzen, Rauschverhalten u. a. unterscheiden können, werden Widerstände unterschiedlicher Herstellungsverfahren verwendet.

Drahtwiderstände werden auf zylindrische Keramikkörper gewickelt. Man erzielt mit ihnen eine hohe Belastbarkeit

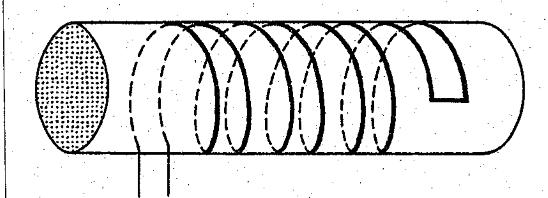


Abb. 8.1.1–1 Bifilar gewickelter Drahtwiderstand

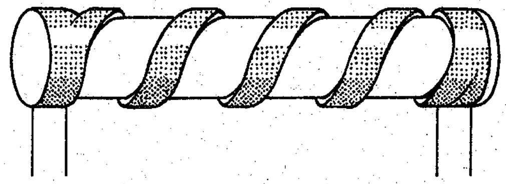


Abb. 8.1.1–2 Schichtwiderstand mit eingefräster Wendel

Tab. 8.1.1–3 Internationaler Farbcode zur Darstellung von Widerstandswerten kleiner Schichtwiderstände

Farbe	1. Marke 1. Ziffer	2. Marke 2. Ziffer	3. Marke Multiplikator	4. Marke Toleranz
keine	–	–	–	±20 %
Silber	–	–	10^{-2}	±10 %
Gold	–	–	10^{-1}	± 5 %
Schwarz	–	0	10^{0}	–
Braun	1	1	10^{1}	± 1 %
Rot	2	2	10^{2}	± 2 %
Orange	3	3	10^{3}	–
Gelb	4	4	10^{4}	–
Grün	5	5	10^{5}	–
Blau	6	6	10^{6}	–
Violett	7	7	10^{7}	–
Grau	8	8	10^{8}	–
Weiß	9	9	10^{9}	–

auf Kosten eines relativ großen Volumenbedarfs. Um die Wicklung induktivitätsarm zu halten, führt man sie oft in bifilarer Form aus (Abb. 8.1.1–1).

Schichtwiderstände haben eine auf einem zylindrischen Keramikkörper aufgetragene, dünne leitende Schicht aus Glanzkohle oder aus mit Bindemitteln versehenen pulverisierten Stoffen, die durch Aufdampfen, Eintauchen oder andere Abscheidungsverfahren auf den Träger aufgebracht werden. Der einzustellende Widerstandswert kann durch Schichtmaterial und -dicke und durch das Einfräsen wendelförmiger Konturen in weiten Grenzen variiert werden (Abb. 8.1.1–2). Auf den Schichtwiderständen kleiner räumlicher Ausdehnung werden Widerstandsnennwert und Toleranzklasse nicht in aufgedruckten Ziffern, sondern durch einen international eingeführten Farbkodex wiedergegeben (Tab. 8.1.1–3).

Massewiderstände bestehen aus einem Pulvergemisch leitender und nichtleitender Substanzen, das in eine zylindrische Form gepreßt wird. In den Kennwerten für Altersbeständigkeit, Rauschverhalten und Temperaturkoeffizient sind sie anderen Widerstandstypen unterlegen und bleiben speziellen Anwendungen vorbehalten.

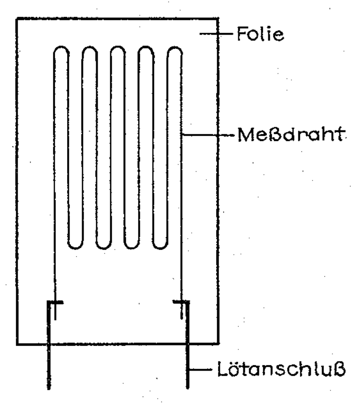
Verstellbare Widerstände können gegenüber den Festwiderständen in Schiebe- und Drehwiderstände eingeteilt werden. Der genutzte Teil des Gesamtwiderstandes wird durch die Stellung eines beweglichen Schleifers bestimmt. Dabei gibt es sowohl verstellbare Widerstände, die ständig veränderbar sein müssen (z. B. bei der Lautstärkeeinstellung eines Rundfunkgerätes), wie auch Widerstände zum einmaligen Abgleich (Trimmer). Schiebe- und Drehwiderstände können je nach Verwendungszweck und Widerstandswert aus Widerstandsdraht oder aus einer Kohleschicht auf Hartpapierträger bestehen. Beim Drehwiderstand z. B. lassen sich zwischen Drehwinkel und dazugehörigem Wider-

standswert verschiedenartige funktionale Abhängigkeit, z. B. linear oder logarithmisch, verwirklichen.

Integrierte Widerstände werden neben diesen als diskrete Bauelemente vorliegenden Widerständen in Festkörperschaltkreisen (vgl. 8.2.) mit vorgegebenem Widerstandswert realisiert. In den mikroelektronischen Technologien, die der Halbleiterblocktechnik vorangegangen sind, der Dickschichttechnologie (Schichtdicken ≈25 µm) und der Dünnschichttechnologie (Schichtdicken ≈0,5 µm) werden Widerstände als linien- oder mäanderförmige Gebiete mit Streifenbreiten um 1 mm ausgebildet. In der Dickschichttechnik erfolgt die Herstellung mit speziellen Pasten im Siebdruckverfahren, während in der Dünnschichttechnik Vakuumaufdampfung oder Katodenzerstäubung bei Benutzung von Masken angewandt werden.

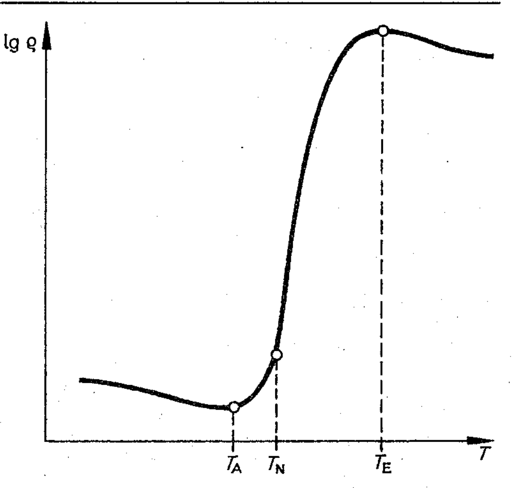
In nach der Halbleiterblocktechnik hergestellten *integrierten Schaltkreisen* können Widerstände auf unterschiedliche Weise verwirklicht werden. Mit Hilfe einer geeigneten Dotierung durch Diffusion oder Ionenimplantation sind einerseits linienhafte Widerstandsstreifen herstellbar, die sich auf der Grundlage geeignet angebrachter *pn*-Übergänge vom benachbarten Halbleitervolumen abgrenzen lassen. Andererseits werden Widerstände auch dadurch realisiert, daß die gewünschte Leitfähigkeit zwischen zwei Punkten über die mit Hilfe der Influenz gesteuerten Ladungsträgerkonzentration eines leitenden Kanals eingestellt wird. Diese Anwendung des Feldeffektes liegt auch dem MIS-Feldeffekttransistor (vgl. 8.1.3.4.) zugrunde.

Für bestimmte Anwendungen werden auch Widerstände benötigt, die ihren Widerstandswert in Abhängigkeit von äußeren Parametern ändern. Dazu gehören die Temperatur (Heißleiter, Kaltleiter), das Magnetfeld (Hallgeneratoren, Magnetwiderstände [sogenannte Feldplatten]), die Beleuchtungsstärke (Fotowiderstand vgl. 8.1.6.1.), die elastische Dehnung des Widerstandsmaterials unter dem Einfluß einer äußeren Kraft (Dehnungsmeßstreifen) und die am Widerstand anliegende Spannung (spannungsabhängige Widerstände [Varistoren]).



8.1.1–4 Schematische Darstellung eines Dehnungsmeßstreifens

Dehnungsmeßstreifen werden zur experimentellen Bestimmung von kleinen, zeitlich veränderlichen Dehnungen bzw. Abstandsänderungen benutzt. Sie bestehen aus dünnem, mäanderförmig auf eine Unterlage aufgeklebten Widerstandsdraht, aufgedampften leitenden Schichten oder aus Halbleiterwiderständen. Auf die Meßstelle aufgeklebt, ist



8.1.1–5 Abhängigkeit des spezifischen Widerstandes eines Bariumtitanat-Keramik-Kaltleiters von der Temperatur (T_A = Beginn von $\alpha_\varrho > 0$, T_N = ferroelektrische Curie-Temperatur, $R(T_N) \approx 2R(T_A)$, Beginn von α_ϱ = const., T_E = Ende von $\alpha_\varrho < 0$)

der auch in seiner zeitlichen Veränderung erfaßbare Dehnungszustand durch die sich am Dehnungsmeßstreifen ergebende Veränderung des elektrischen Widerstandes charakterisierbar und kann z. B. mit einem Katodenstrahloszillografen sichtbar gemacht werden (Abb. 8.1.1–4).

Als Maß für die Empfindlichkeit k eines Dehnungsmeßstreifens wählt man das Verhältnis der relativen Widerstandsänderung $\Delta R/R$ zur relativen Längenänderung $\Delta l/l = \varepsilon$.

Mit der Einführung der Poissonschen Querkontraktionszahl $\nu = -(\Delta r/r)/(\Delta l/l)$, die das Verhältnis von Querkontraktion $-\Delta r/r$ zur Längsdehnung $\varepsilon = \Delta l/l$ beschreibt, erhält man für $k = (1/\varepsilon)(\Delta R/R), 1 + 2\nu + (1/\varepsilon)(\Delta\varrho/\varrho)$ (ϱ = spezifischer elektrischer Widerstand).

Der Wertebereich der Poisson-Zahl liegt zwischen $0 < \nu < 0{,}5$. Für metallische Dehnungsmeßstreifen ist die relative Änderung des spezifischen Widerstands vernachlässigbar, so daß für die Empfindlichkeit metallischer Dehnungsmeßstreifen $1 < k < 2$ gilt.

Für Halbleiter-Dehnungsmeßstreifen liefert die relative Änderung des spezifischen Widerstands den Hauptbeitrag für k, das für Silicium-Dehnungselemente auf Werte um 100 ansteigt. Sie werden auch als Piezowiderstände bezeichnet. Mit ihnen lassen sich relative Längenänderungen bis herab zu 10^{-3} bequem nachweisen. Dehnungsmessungen sind für die Erfassung des zeitlich periodischen Verlaufes besonders an vibrierenden Konstruktionen, wie etwa an Flugzeugteilen, Brücken, rotierenden Maschinenteilen oder an Rohrleitungen, wichtig. Durch geeignete Anordnung mehrerer Dehnungsmeßstreifen kann man einerseits Temperatureinflüsse eliminieren, oder andererseits auch Drehmomente, etwa an rotierenden Achsen, messen.

Wegen der Linearität zwischen relativer Dehnung und relativer Widerstandsänderung verwendet man Halbleiter-Dehnungsmeßstreifen unter Einbeziehung von Kompensations- und Verstärkerbauelementen auch als integrierte Bauelemente, die aus meßtechnischen Gründen zusätzlich mit einem Oszillator ausgestattet sein können. Anordnungen

dieser Art eignen sich zur Druckmessung zwischen 10^2 Pa und 10 MPa und können vielseitig, z. B. für die Blutdruckmessung in der Medizin oder die Messung des Staudrucks in der Aerodynamik, eingesetzt werden.

Kaltleiter sind Bauelemente, deren Widerstand mit sinkender Temperatur abnimmt, ihr Temperaturkoeffizient des spezifischen Widerstands $|\alpha_\varrho| = (1/\varrho)(\mathrm{d}\varrho/\mathrm{d}t)$ ist positiv und selbst eine Funktion der Temperatur (PTC-Widerstände). Zu den Kaltleitern gehören auch die Metalle, die bei konstanter Temperatur einen linearen Zusammenhang zwischen Strom und Spannung besitzen und deren Widerstand sich innerhalb eines Temperaturintervalls von $\Delta T = 100$ K etwa um die Hälfte vergrößert ($\alpha_\varrho \approx 5 \cdot 10^{-3}\,\mathrm{K}^{-1}$).

Innerhalb eines Temperaturintervalls, das je nach der Materialzusammensetzung in bestimmten Grenzen wählbar ist, hat Bariumtitanatkeramik einen sehr großen positiven Temperaturkoeffizienten α_ϱ des spezifischen Widerstands, der zwischen $0{,}1\,\mathrm{K}^{-1}$ und $0{,}3\,\mathrm{K}^{-1}$ liegen kann (im Vergleich zu $\alpha_\varrho = 4{,}33 \cdot 10^{-3}\,\mathrm{K}^{-1}$ für Kupfer) und eine stark nichtlineare Strom-Spannungs-Kennlinie. Außerhalb dieses Intervalls ist der Temperaturkoeffizient des spezifischen Widerstands $\alpha_\varrho < 0$ (Abb. 8.1.1–5).

Die Wirkung des Bariumtitanats als Kaltleiter tritt nur bei polykristallinem Material auf. Für diesen Sachverhalt werden die an den Korngrenzen auftretenden Übergangswiderstände verantwortlich gemacht. Die Temperatur für den Einsatz des Widerstandsanstiegs (vgl. Abb. 8.1.1–5) ist durch Substitution des Bariums bzw. des Titaniums in weiten Grenzen veränderbar. Die Bauelemente haben meist Scheibenform und an das keramische Material angelötete Zuleitungsdrähte. Zum Schutz vor atmosphärischen Einflüssen sind die Bauelemente mit einem Lacküberzug versehen.

Wenn Kaltleiter als fremderwärmte Bauelemente eingesetzt werden, kann mit ihnen auf elektrischem Wege die Temperatur gemessen werden. Dies wird insbesondere bei Schutz von Elektromotoren vor Überhitzung genutzt.

Bei der Erwärmung von Kaltleitern durch die in ihnen erzeugte Stromwärme bestimmt die Wärmeableitung an die Umgebung den sich einstellenden Widerstandswert. Dieser Umstand wird zur Strömungsmessung in gasförmigen Medien und zur Füllstandsanzeige in Flüssigkeitsbehältern angewandt. Da die durch einen großen Strom einsetzende Erwärmung des Kaltleiters mit einer zeitlichen Verzögerung verbunden ist, eignen sich Kaltleiter auch dazu, z. B. den Selbstanlauf von Einphasen-Induktionsmotoren herbeizuführen.

Hierzu ist ein genügend hoher Strom durch eine Hilfswicklung erforderlich, der nach erfolgtem Anlauf wieder abgeschaltet werden muß.

Heißleiter sind temperaturabhängige Widerstände aus Halbleitermaterialien mit, gemessen am Betrag des Temperaturkoeffizienten der Metalle, vergleichsweise großem negativem Temperaturkoeffizienten des elektrischen Widerstands (NTC-Widerstände). Sie werden als Einzelbauelemente in Perlen-, Stäbchen- oder Scheibenform gepreßt und bei Temperaturen oberhalb 1 300 K gesintert. Als keramische Heißleiter bestehen sie aus Oxiden der Übergangsmetalle Mangan, Eisen, Cobalt, Nickel wie auch von Kupfer und Zink, die sich teilweise zu Spinellen umbilden.

Bei Erwärmung werden zusätzliche freie Ladungsträger erzeugt und die Ladungsträgerbeweglichkeit ändert sich.

Komponentenart, Mischungsverhältnis und Herstellungstechnologie lassen eine vielfältige Parametervariation zu.

Die Abhängigkeit des Widerstands von der Temperatur läßt sich durch die Beziehung $R(T) = R_\mathrm{N} \exp\{B(1/T - 1/T_\mathrm{N})\}$ hinreichend genau darstellen. B ist hier eine Stoffkonstante und R_N stellt den Widerstand bei einer Bezugstemperatur T_N, die meist bei 300 K gewählt wird, dar. R_N hängt von der Bauform und der Art des Werkstoffes ab. Der Temperaturkoeffizient des elektrischen Widerstands eines Heißleiters $\alpha_R = (1/R)(\mathrm{d}R/\mathrm{d}T) = -B/T^2$ ist selbst stark temperaturabhängig.

Für die Anwendung kann – wie bei Kaltleitern – die Wärmeenergie entweder von außen einwirken (Fremderwärmung) oder von einem durch das Bauelement fließenden Strom erzeugt werden (Eigenerwärmung).

Die stark nichtlineare Widerstands-Temperatur-Kennlinie der Heißleiter läßt sich z. B. für den Zweck einer Temperaturmessung auf elektrischem Wege durch Parallelschaltung eines Festwiderstands linearisieren. Die Linearisierung reduziert allerdings den Wert des Temperaturkoeffizienten des elektrischen Widerstands der Kombination.

In Brückenschaltung lassen sich Temperaturdifferenzen messen. Wie auch Kaltleiter, eignen sich Heißleiter zur Messung von Strömungsgeschwindigkeiten und zur Füllstandsanzeige. Die thermische Trägheit von Heißleitern eröffnet ebenfalls eine Reihe von Anwendungsmöglichkeiten, von denen die Anzugs- bzw. Abfallverzögerung von Relais genannt werden soll. Wenn ein Heißleiter mit der Erregerwicklung eines Relais in Reihe liegt, so spricht das Relais erst dann an, wenn sich der Heißleiter so erwärmt hat, daß der zum Anziehen des Relais nötige Strom fließt.

Daneben nutzt man Heißleiter auch, um Temperaturabhängigkeiten anderer Art zu kompensieren sowie zur Spannungsstabilisierung. Den zur Spannungsstabilisierung nötigen nahezu horizontalen Verlauf der Strom-Spannungs-Kennlinie erhält man, falls er nicht schon durch den Heißleiter selbst gegeben ist, durch Reihenschaltung mit einem Festwiderstand (Abb. 8.1.1–6).

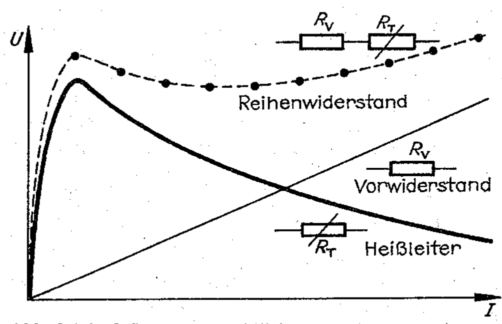


Abb. 8.1.1–6 Spannungsstabilisierung mit Heißleiter und Reihenwiderstand

Galvanomagnetische Bauelemente sind Widerstände, die ihren Wert unter dem Einfluß eines magnetischen Feldes ändern. Zu ihnen zählen das Hall-Element und die Feldplatte. Beide nutzen den Umstand, daß auf einen bewegten Ladungsträger im Magnetfeld zwischen den Zusammenstößen mit den Gitterbausteinen die Lorentz-Kraft $\boldsymbol{F}_\mathrm{L} = e[\boldsymbol{v} \times \boldsymbol{B}]$ wirkt, e = Betrag der Elementarladung, $\boldsymbol{v}$ = Geschwindigkeit des Ladungsträgers, $\boldsymbol{B}$ = magnetische Induktion.

Hall-Elemente sind entsprechend der Anordnung zum Nachweis des Hall-Effektes aufgebaut (Abb. 8.1.1–7). Sie haben die Form eines Plättchens, bestehen aus der $\mathrm{A^{III}B^{V}}$-

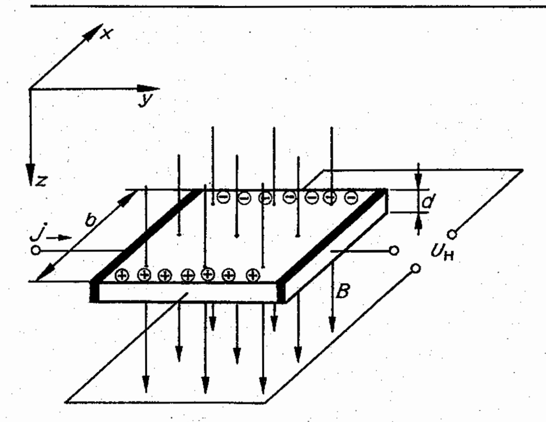


Abb. 8.1.1–7 Schema eines Hall-Elementes

Verbindung Indiumantimonid- oder -arsenid und sind an den Längsseiten mit flächenhaften Kontakten zur Zuführung des Primärstroms j und an den Seitenflächen mit punktförmigen Kontakten zur Abnahme der durch das Magnetfeld verursachten Hall-Spannung U_H versehen.

Der makroskopische Hall-Winkel Φ_H beschreibt die Verbiegung der elektrischen Feldlinien unter dem Einfluß des Magnetfeldes. Es ist $\tan \Phi_H = E_x/E_y = -\mu B$. Die direkte Proportionalität zu μ erklärt, warum sich InSb und InAs besonders für Hall-Elemente eignen. Ihre Elektronenbeweglichkeiten liegen mit $\mu_n = 78\,000\ \mathrm{cm^2/Vs}$ bei InSb und mit $\mu_n = 33\,000\ \mathrm{cm^2/Vs}$ bei InAs sehr hoch.

Mit einem Magnetfeld von $B = 1$ Tesla stellt sich der Hallwinkel Φ_H für InSb etwa auf 80° ein, während die gleiche Feldstärke bei Metallen nur zu Werten von $\Phi_H < 0{,}5°$ führen würde. Hall-Elemente werden aus kristallinem Material oder auch aus Aufdampfschichten hergestellt. In integrierter Form gibt es auch Hall-Elemente aus Silicium, die zusätzlich mit Verstärkeranordnungen kombiniert sind.

Die zwischen den Punktkontakten an den Seitenflächen vorhandene Hall-Spannung U_H ergibt sich zu $U_H = E_H b = R_H jB$, R_H = Hall-Konstante, E_H = Hall-Feldstärke, b = Breite der Flächen. Da die Hall-Spannung dem Produkt der beiden Steuergrößen Primärstrom und Magnetfeld B proportional ist, kann das Hall-Element die Multiplikation beider Größen ausführen.

Hall-Elemente werden zur Messung von Magnetfeldern sowie bei deren Regelung oder Konstanthaltung, aber auch z. B. bei der Abtastung magnetisierbarer Folien eingesetzt. Zur Charakterisierung von Leiter- oder Halbleitermaterialien gehören solche wichtige Kenngrößen, wie Ladungsträgerbeweglichkeit μ und -konzentration n. Durch Messung der Leitfähigkeit erhält man entsprechend $\sigma = en\mu$ nur das Produkt beider. Bringt man die zu charakterisierende Substanz in die Form eines Hall-Elementes, so kann aus der Messung der Hall-Spannung die Ladungsträgerkonzentration n bestimmt werden. Diese Größe ist in der Hall-Konstanten R_H enthalten, da $E_x = E_H = E_y \tan \Phi_H = -\mu B E_y = 1/(en)jB$, womit $R_H = 1/(en)$ wird.

Die Hall-Konstante R_H enthält auch das Ladungsträgervorzeichen, so daß zwischen Elektronen- und Defektelektronenleitung unterschieden werden kann.

Feldplatten beruhen auf der Widerstandsänderung im Magnetfeld. Während beim Hall-Element die sich quer zur Primärstromrichtung unter dem Einfluß des senkrecht verlaufenden Magnetfeldes einstellende Hall-Spannung genutzt wird, interessiert man sich bei der Feldplatte direkt für die unter dem Einfluß des Magnetfeldes auftretende Strom- bzw. Widerstandsänderung. Es ergibt sich unter Zuhilfenahme der Interpretation des Hall-Effektes die relative Widerstandsänderung im transversalen Magnetfeld von der Größenordnung $\Delta\varrho/\varrho = (\mu B)^2$. Es handelt sich um einen quadratischen Effekt des Magnetfeldes, der um so stärker ausfällt, je größer die Ladungsträgerbeweglichkeit μ ist.

Bei der Anwendung der Feldplatte muß auf die Geometrie des Halbleiterplättchens geachtet werden. Die größte Widerstandsänderung liefert die kreisförmige Corbino-Scheibe, bei der eine Elektrode den Kreisumfang, die andere den Kreismittelpunkt bildet.

Rechteckige Plättchen mit einem kleinen Längen-Breiten-Verhältnis erzielen große Widerstandsänderungen. Zweckmäßig wäre also die Reihenschaltung vieler solcher Plättchen zu einer Rasterplatte, in der die Plättchen durch leitende Stege miteinander verbunden sind. Technologisch einfacher ist der orientierte Einbau von gut leitenden NiSb-Nadeln quer zur Stromrichtung, die dieselbe Wirkung wie die leitenden Stege einer Rasterplatte ergeben. Die Widerstandsbahn kann auch mäanderförmig ausgeführt sein.

Feldplatten dienen ebenfalls der Messung, Steuerung und Regelung von Magnetfeldern. Sie lassen sich als kontaktlos und durch Elektromagneten oder bewegliche Dauermagneten kontinuierlich steuerbare Widerstände oder kontaktlose Potentiometer verwenden.

Varistoren sind spannungsabhängige Widerstände mit nichtlinearer Strom-Spannungs-Kennlinie (Abb. 8.1.1–8). Sie bestehen aus kleinen SiC-, TiO_2- oder ZnO-Körnern, die – mit einem Bindemittel versehen – zwischen 1 000 und 1 300 °C gesintert und zwischen metallische Elektroden gepreßt werden. Da sich bei Spannungsänderungen der zugehörige Widerstand nahezu momentan einstellt, ist der Varistoreffekt den Grenzschichten zwischen den Kristalliten zuzuschreiben, die sich diodenartig verhalten. Der Verlauf der Strom-Spannungs-Kennlinie eines Varistors stimmt demzufolge auch gut mit der Strom-Spannungs-Kennlinie eines Komplexes von parallel und in Reihe geschalteten Dioden überein.

Die Strom-Spannungs-Kennlinie von Varistoren läßt sich durch die Beziehung $U = CJ^{\beta}$ annähern. β bestimmt das Maß der Nichtlinearität der Kennlinie und ist ein Substanzparameter. Es gilt $0{,}02 \leqq \beta \leqq 0{,}4$. C ist ein Formparameter und durch das Länge-Durchmesser-Verhältnis des Bauelementes bestimmt. Die Größe C liegt zwischen 100 und 3 000. Varistoren werden in Spannungsteilern mit einem Ohmschen Widerstand zur Spannungsstabilisierung

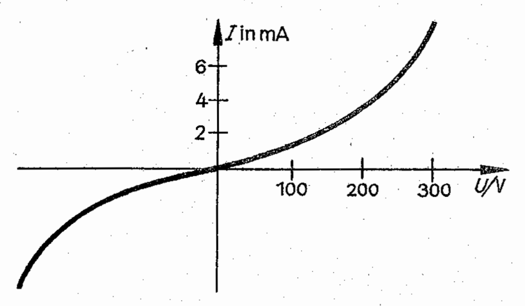


Abb. 8.1.1–8 Strom-Spannungs-Kennlinie eines Varistors (linearer Maßstab)

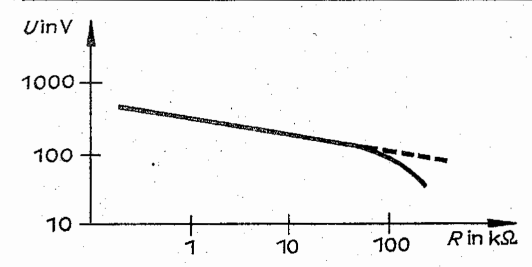


Abb. 8.1.1–9 Abhängigkeit des Widerstandes von der Spannung bei einem Varistor

benutzt (Abb. 8.1.1–9). In Entzerrerschaltungen können sie z. B. zur Linearisierung des Anzeigefeldes von Meßinstrumenten dienen. Bei abzuschaltenden Induktivitäten verhindern parallelgeschaltete Varistoren eine Funkenbildung.

8.1.1.2. Kondensatoren bestehen aus zwei voneinander durch ein Dielektrikum getrennten Metallelektroden. Die wichtigste Kenngröße eines Kondensators, die Kapazität C, läßt sich für beliebige Elektrodenanordnungen in Dielektrika berechnen. Von praktischer Bedeutung sind besonders platten- und zylinderförmige Kondensatoren.

Plattenkondensator. Bei ihm ist Q die auf einer Platte befindliche Ladung. Die Größe U kann als Potential dieser Kondensatorfläche bezüglich der gegenüberliegenden Fläche aufgefaßt werden. Für die Kapazität C des Plattenkondensators erhält man $C = Q/U = Q/(Ed) = Q/(Qd/[\varepsilon A]) = \varepsilon A/d$ (A = Fläche einer Kondensatorplatte, d = Plattenabstand, $\varepsilon = \varepsilon_0 \varepsilon_r$, ε_0 = elektrische Feldkonstante, $\varepsilon_0 = 8\,854 \cdot 10^{-12}$ As/Vm , ε_r = Dielektrizitätszahl, im Vakuum gilt $\varepsilon_r = 1$, ε = Dielektrizitätskonstante).

Die Kapazität wächst mit zunehmender Plattenfläche, zunehmender Dielektrizitätskonstante und abnehmendem Plattenabstand. Somit wird auch verständlich, daß sich bei Parallelschaltung von Kondensatoren (Vergrößerung der Flächen) deren Kapazitäten zur Gesamtkapazität addieren, $C = C_1 + C_2$.

Bei Reihenschaltung von Kondensatoren addieren sich dagegen die Kehrwerte der Kapazitäten zum Kehrwert der Gesamtkapazität $1/C = 1/C_1 + 1/C_2$.

Zylinderkondensatoren (Abb. 8.1.1–10) bestehen aus einem dielektrischen Röhrchen mit dem Innenradius r und dem Außenradius R. Innen- und Außenfläche sind mit einem leitenden Belag versehen. Die Kapazität des Zylinderkondensators berechnet sich zu $C = 2\pi\varepsilon l/(\ln[R/r])$ (l = Zylinderlänge). Legt man eine Gleichspannung an einen Kondensator, so erfolgt in einem Stromstoß die Aufladung der Kondensatorplatten. Im weiteren Verlauf stellt er (im Idealfall) einen unendlich großen Widerstand dar. Eine Wechselspannung an den Kondensatorelektroden bewirkt eine ständige Umladung der Platten. Es kommt ein Strom zustande, der umso größer wird, je häufiger der Wechsel erfolgt, d. h. je höher die Kreisfrequenz ω und je größer die Kapazität C des Kondensators ist.

Der kapazitive Widerstand $X_C = 1/(\omega C)$ ist ein Blindwiderstand, weil der zeitliche Mittelwert der Leistung am Kondensator verschwindet.

Im Vergleich zu den Eigenschaften eines idealen Kondensators haben technische Kondensatoren außer der gewünschten Kapazität je nach Bauart und Technologie der Herstellung noch weitere elektrische Eigenschaften, die bei der Anwendung zu berücksichtigen sind. Die wichtigsten sind der über das Dielektrikum fließende Reststrom, die Durchschlagsfestigkeit und der temperatur- und frequenzabhängige Verlustfaktor $\tan\delta$. Der Verlustwinkel bezeichnet den Winkel, um den die Phasenverschiebung kleiner als $\pi/2$ ist.

Man unterscheidet zwischen den frequenzunabhängigen Isolationsverlusten infolge des endlichen Isolationswiderstands des Dielektrikums, den stoff- und frequenzabhängigen dielektrischen Verlusten infolge der elastischen Verformung der Moleküle des Dielektrikums und den erst bei höheren Frequenzen wirksam werdenden Nachwirkungsverlusten infolge des zeitlichen Zurückbleibens der Polarisation gegenüber der elektrischen Feldstärke. Die präzise Beschreibung der Eigenschaften technisch realisierter Kondensatoren fordert somit die Verwendung eines Ersatzschaltbildes. In diesem Ersatzschaltbild liegt parallel zur Kapazität des Kondensators ein Ohmscher Widerstand, der den Grad der Isolation zwischen den Elektroden berücksichtigt. Dieser Parallelschaltung folgt ein Serienwiderstand, der Zuleitungs- und Übergangswiderstände enthält sowie eine Serieninduktivität, die die Induktivität des Kondensatorkörpers, des Wickels und der Anschlüsse enthält. Insgesamt lassen sich folgende Eigenschaften von Kondensatoren in elektrischen Netzwerken aufführen: Sie stellen Blindwiderstände dar, bewirken eine Phasenverschiebung zwischen Strom und Spannung mit nacheilender Spannung, sie speichern elektrische Energie, sie sind für Gleichströme unendlich große Widerstände, für Wechselströme ist ihr Widerstandswert der Frequenz des Stroms und der Kapazität umgekehrt proportional. Diese Eigenschaften werden mit Hilfe von Kondensatoren verschiedener Bauart für vielfältige Anwendungen genutzt. Einige typische Beispiele seien angeführt: Kondensatorbatterien hoher Kapazität dienen als Speicher für kurzzeitige Stromstöße extrem hoher Stromstärken, in kleineren Ausführungen können Blitzlichtgeräte gezündet werden. Kondensatoren dienen als frequenzbestimmende Bauelemente in Schwingkreisen. Sie koppeln bei gleichspannungsmäßiger Trennung aufeinanderfolgende Verstärkerstufen aneinander. Sie dienen in Siebketten zur Glättung von gleichgerichtetem Wechselstrom. Als Überbrückung verhindern sie als Störschutzkondensatoren Entstehung und Abstrahlung hochfrequenter Störungen. In Starkstromnetzen nutzt man sie als Phasenschieber-Kondensatoren zur Verbesserung des Leistungs-

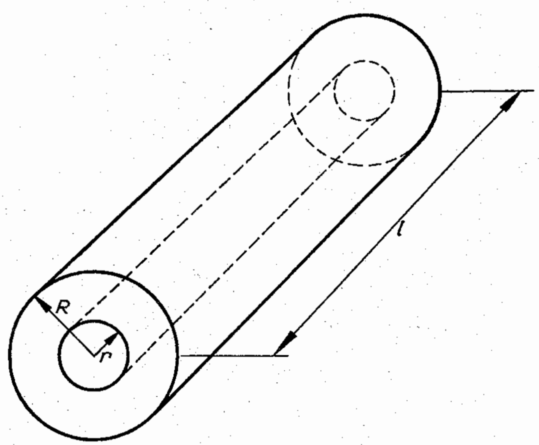


Abb. 8.1.1–10 Zylinderkondensator

faktors cos φ. Entsprechend dem Einsatzgebiet und den erforderlichen Eigenschaften sind eine Reihe von Technologien der Herstellung von Kondensatoren üblich, von denen einige angeführt werden.

Papierkondensatoren bestehen aus zwei Aluminiumfolien (Foliendicken $\approx 10\,\mu m$) mit ein- oder mehrlagigen, ölgetränkten Papierbändern als Dielektrikum. Mit angeschweißten Anschlußdrähten wird die Anordnung aufgewickelt und in Metallbecher, Hartpapier- oder Keramikröhren eingesetzt. Sie werden mit Kapazitäten von ≈ 100 pF bis zu $\approx 100\,\mu$F hergestellt und beispielsweise in der Fernsehtechnik wegen ihrer Spannungsfestigkeit genutzt.

Metallpapierkondensatoren bestehen aus einem $\approx 10\,\mu m$ dicken Papierstreifen, auf den eine $\approx 0{,}1\,\mu m$ dicke Aluminiumschicht aufgedampft wird. Als Becher- oder Rohrkondensatoren erzielt man bei gleichen Volumina beträchtliche Kapazitätserhöhungen. Bemerkenswert ist die Eigenschaft der Selbstausheilung bei elektrischem Durchschlag. Die lokale Verdampfung der Metallschicht verhindert einen Kurzschluß zwischen den Belägen. Metallpapierkondensatoren werden mit Kapazitätswerten von ≈ 50 nF bis zu einigen 10 µF hergestellt. Ihre hohe Alterungsbeständigkeit empfiehlt sie besonders als Glättungskondensatoren.

Kunststoffoliekondensatoren. Anstelle des Papiers wird eine Kunststoffolie (z. B. Polystyren, Polyester, Lackfilm) verwendet, auf die die Metallbeläge wie bei Metallpapierkondensatoren aufgedampft oder aufgespritzt sind. Die Festigkeit der Folien und die im Vergleich zum Papier bessere Homogenität ermöglicht einlagige Folien. Trotz der kleineren Dielektrizitätskonstante im Vergleich zu Papier lassen sich je Volumeneinheit höhere Kapazitätswerte erzielen.

Keramikkondensatoren unterscheidet man nach den dielektrischen Eigenschaften der verwendeten Ausgangsmaterialien, die in Scheiben oder Röhrchen geformt und anschließend gesintert werden. Keramische Kondensatoren mit kleiner Dielektrizitätskonstante (NDK-Kondensatoren) enthalten Titaniumdioxid (TiO_2) in Form von Rutil als Ausgangswerkstoff. Durch geeignete Zuschlagsstoffe wird der Temperaturkoeffizient der Dielektrizitätskonstante beeinflußt.

Keramische Kondensatoren mit hoher Dielektrizitätskonstante (HDK-Kondensatoren), d. h. $2\,000 \leqq \varepsilon \leqq 10\,000$, enthalten das Ferroelektrikum Bariumtitanat ($BaTiO_3$) als Ausgangsstoff. Kapazitätswerte von NDK-Kondensatoren liegen zwischen 1 pF und 1 nF, die von HDK-Kondensatoren je nach Bauart zwischen 100 pF und 100 nF. NDK-Kondensatoren verwendet man bei Forderung hoher Stabilität, kleiner Verluste und vorgegebenem Temperaturverhalten in Hochfrequenzkreisen und -filtern, während HDK-Kondensatoren bei gefordertem hohen Kapazitäts-Volumen-Verhältnis eingesetzt werden.

Elektrolytkondensatoren dürfen aufgrund ihres Aufbaus nur in festgelegter Polung betrieben werden. Die Anode ist eine $\approx 50\,\mu m$ dicke Aluminiumfolie. Das Dielektrikum wird durch eine Oxidhaut von Aluminiumoxid, Al_2O_3, ($\varepsilon \approx 8$) gebildet, die je Volt Betriebsspannung eine Dicke von mindestens 1,1 bis 1,4 nm aufweisen muß. Die so aufgewickelte Aluminiumfolie ist mit einer leitenden Flüssigkeit umgeben, die auf elektrolytischem Wege Sauerstoff abgeben kann und bei geeigneter Polung die Oxidschicht aufrecht erhält. Die erstmalige Erzeugung der Oxidhaut auf elektrochemischem Wege bei der Herstellung des Kondensators bezeichnet man als Formierung. Die Kontaktierung des Elektrolyten, der Katode, erfolgt durch eine Metallfolie oder durch das Kondensatorgehäuse selbst. Eine Aufrauhung der Aluminiumfolie vor der Formierung vergrößert die wirksame Plattenfläche und kann die Kapazität auf ein Vielfaches erhöhen. Bei falscher Polung führt ein hoher Flußstrom zum schnellen Abbau der Oxidschicht und damit zur Zerstörung des Elektrolytkondensators. Da die Durchschlagfestigkeit von der Dicke der Oxidschicht abhängt, lassen sich bei gleichen Volumina für niedrige Spannungen Elektrolytkondensatoren mit den größten Kapazitäten herstellen. Verwendet man anstelle von Aluminium Tantal, so kann man die Anode anstatt folienartig als porösen Sinterkörper ausbilden. Das Dielektrikum ($\varepsilon = 27$) in Form einer Tantalpentoxidschicht (Ta_2O_5) mit einer erforderlichen Dicke von 1,7 bis 2,1 nm je Volt Betriebsspannung läßt sich entweder mit dem trockenen Mangandioxid (MnO_2) oder bei nassen Tantal-Elektrolytkondensatoren durch Schwefelsäure formieren und aufrechterhalten. Tantal-Elektrolytkondensatoren haben einen kleineren Reststrom und wesentlich kleinere Abmessungen als Aluminium-Elektrolytkondensatoren gleicher Kapazität.

Drehkondensatoren sind Kondensatoren variabler Kapazität. Man benötigt sie insbesondere zur Abstimmung von Schwingkreisen. Die technische Realisierung erfolgt durch zwei relativ zueinander verdrehbare, kammartig ineinandergreifende Plattenpakete. Dabei bewegt sich das eine Plattenpaket in den Zwischenräumen des anderen. Entsprechend dem ausgebildeten Plattenprofil ändern Drehkondensatoren ihre Kapazität logarithmisch oder linear mit dem Drehwinkel.

Trimmer sind Drehkondensatoren, die nur zu einem einmaligen Abgleich veränderlich sein müssen.

In integrierten Schaltkreisen lassen sich Kondensatoranordnungen mit vertretbaren Flächengrößen nur bis zu einigen 100 pF verwirklichen. In der Bipolartechnik bietet sich die sonst nicht erwünschte Grenzschichtkapazität eines in Sperrichtung gepolten *pn*-Übergangs als Kondensator an, wobei für die maximalen Kapazitätswerte bereits Flächen von 1 mm² erforderlich sind. In der MIS-Technik werden integrierte Kondensatoren aus der Kapazität zwischen einem hochdotierten Kanalgebiet und einer Metallelektrode gebildet. Die dazwischenliegende SiO_2-Isolationsschicht übernimmt die Funktion des Dielektrikums. Diese Lösung vermeidet viele Nachteile der Sperrschichtkapazitäten.

8.1.1.3. Spulen können als Serienschaltung koaxialer Leiterschleifen aufgefaßt werden.

In verlustfreien Spulen (d. h. u. a., wenn der Ohmsche Widerstand der Leiterschleifen vernachlässigt werden kann) wird keine Energie in Wärme umgewandelt. Damit sind Spulen einerseits zur Erzeugung von Magnetfeldern verwendbar (*Erregerspulen*), die bestimmte Wirkungen, beispielsweise die Schließung von Relaiskontakten zu bewirken haben.

Andererseits dienen Spulen als *Induktionsspulen* bei Nutzung der Selbstinduktion als Drosselspulen und bei Nutzung der Gegeninduktion als Transformator, Übertrager bzw. Spannungswandler.

Wie jeder stromführende Leiter, baut eine Spule in ihrer Umgebung ein Magnetfeld auf.

Der Zusammenhang zwischen magnetischem Fluß Φ und der Stromstärke I im Leiter ist gegeben durch $\Phi = L \cdot I$. Der Proportionalitätsfaktor L heißt Selbstinduktionskoeffizient oder Induktivität der Spule.

In Analogie zum Ohmschen Gesetz hat man für einen Wechselstromkreis mit Selbstinduktion L einen Widerstand vom Betrag $X_L = \omega L$, den man als induktiven Widerstand bezeichnet.

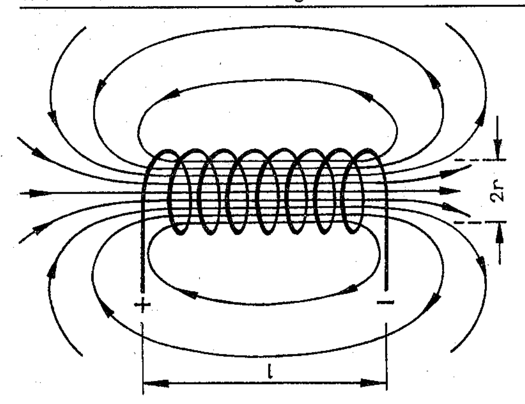


Abb. 8.1.1–11 Stromdurchflossene Zylinderspule und ihre magnetischen Feldlinien

Für einige wenige Spulenarten läßt sich die Induktivität (genauer der Selbstinduktionskoeffizient) L in einfacher Weise berechnen. Das gilt für die *Zylinderspule* (Abb. 8.1.1–11). Die Spulenfläche sei $A = \pi r^2$ mit $r \ll l$, die Spulenlänge. Die Spule habe n Windungen. Für den Selbstinduktionskoeffizienten L der Zylinderspule folgt unter der gegebenen Näherung $L = \mu\mu_0 n^2 A/l$.
Läßt man die Näherung $r \ll l$ fallen, so ist je nach dem Verhältnis von l/r die Größe L mit einem Zahlenfaktor zu multiplizieren, der für $l/r = 0{,}2$; 2; 20 die Größen 0,2; 0,6; $\approx 1{,}0$ annimmt. Hier zeigt sich der Umstand, daß stromdurchflossene Spulen je nach ihrer Bauart nichtverschwindende Felder im Außenraum aufweisen, was die exakte Berechnung oft erschwert. In vielen Fällen wird deshalb mit Näherungen gearbeitet. Eine einfache Angabe des Selbstinduktionskoeffizienten ist noch für die *Ringspule* (Toroid) möglich, bei der die Spulenlänge l einem Kreisumfang $U = 2\pi R$ entspricht. Die analoge Anwendung des Durchflutungsgesetzes ergibt $L = \mu\mu_0 n^2 A/(2\pi R)$.
Je nach dem vorgesehenen Anwendungszweck sind zahlreiche Wicklungsarten bekannt. Spulen können ein- oder mehrlagig, pyramidenartig oder sich überkreuzend gewikkelt sein. Sie sind entweder freitragend oder befinden sich auf einem Wickelkörper, der aus mehreren Kammern bestehen kann. Das Leitermaterial ist isolierter Kupferdraht und besteht für hohe Frequenzen wegen des Skineffekts aus Litze. Spulen, die nur eine geringe Eigenkapazität haben dürfen, müssen einen vergleichsweise großen Windungsabstand aufweisen. In der technischen Anwendung nutzt man zur Charakterisierung von Spulen den Induktivitätsfaktor $A_L = L/n^2 = 1/R_m$ (R_m = magnetischer Widerstand).
Die meisten Spulen haben Kerne ferromagnetischer Substanzen mit relativen Permeabilitäten $\mu \gg 1$. Dadurch wird der magnetische Fluß Φ, die Flußdichte B im Spuleninnern vergrößert, der magnetische Widerstand R_m verringert. Die ferromagnetischen Kernmaterialien zeigen in der Abhängigkeit der Flußdichte B vom äußeren Magnetfeld H eine Hysteresekurve. Die Induktivität von Spulen mit Eisenkern hängt damit auch von der Gleichstromvormagnetisierung (dem „Arbeitspunkt" auf der Hysteresekurve) ab.
Für Drosselspulen im Niederfrequenzgebiet werden Kerne unterschiedlichen Zuschnitts aus gegeneinander isolierten Blechen verwendet, um die entstehenden Wirbelstromverluste klein zu halten. Höhere Frequenzen verlangen Kerne aus mit einem Bindemittel zusammengepreßten Karbonyleisenpulver bzw. Ferritkerne. Eine wichtige technische Kenngröße ist die Güte Q der Spule, das Verhältnis von Blindleistung P_b zu Wirkleistung P_w, oder der Kehrwert des Tangens des Verlustwinkels δ, jenes Winkels, um den die Phasenverschiebung zwischen Strom und Spannung kleiner als $\pi/2$ ist: $Q = P_b/P_w = 1/\tan\delta$. Die Güte wird für alle realen Spulen durch die in den Wicklungen umgesetzte Stromwärme, eine Wirkleistung, endlich. Für Spulen mit ferromagnetischem Kern kommen Wirkleistungen infolge der erzeugten Wirbelströme hinzu. Einen weiteren Beitrag liefern die Energieverluste infolge der Ummagnetisierung. Ein Maß dafür ist der Flächeninhalt der Hysteresekurve. Schließlich stellt sich die Magnetisierung des Kerns nicht gleichzeitig mit dem Feld ein. Dadurch entstehen die sogenannten Nachwirkungsverluste.
Spulen sind frequenzbestimmte Bestandteile von Schwingkreisen und als solche auch in Filtern enthalten. Als Filter bezeichnet man Netzwerke mit einem vorgegebenen frequenzabhängigen Übertragungsverhalten. Für Überlagerungsempfänger braucht man Zwischenfrequenzbandfilter, deren Durchlaßbereich die gebräuchlichen Zwischenfrequenzen von 456 kHz und 10,7 MHz enthält. Dazu nimmt man zwei meist induktiv gekoppelte Schwingkreise, deren Resonanzfrequenz zur Erzeugung spezieller Durchlaßkurven eng beieinanderliegen. Technologisch gesehen, ist die Herstellung von Spulen relativ aufwendig. Es gibt zahlreiche Versuche, die Wickeltechnik durch flächenhafte Induktivitäten, die z. B. durch Siebdruckverfahren oder Leiterplattentechnologie hergestellt sind, zu ersetzen und diese durch Faltung zu räumlichen Spulen zusammenzufügen.
In die mikroelektronische Halbleiterblocktechnik können Spulen, also Induktivitäten und Gegeninduktivitäten nicht integriert werden. Dieses Erfordernis muß durch spezielle schaltungstechnische Maßnahmen umgangen werden. So werden beispielsweise passive Filter, die aus Spulen und Kondensatoren bestehen, durch aktive RC-Filter ersetzt.
Da stromdurchflossene Spulen auch auf ihren Außenraum wirken, ist eine wechselseitige Beeinflussung zweier Spulen L_1 und L_2 möglich. Der magnetische Fluß einer Spule L_1 durchdringt teilweise die Spule L_2. Eine Stromänderung in L_1 bewirkt dann eine Änderung des magnetischen Flusses in L_2 und damit die Induzierung einer Spannung an ihren Leiterenden. Diese Spannung $u_{\text{ind}2}$ an L_2 wird als Folge einer Gegeninduktivität M_{21} bezüglich L_1 gedeutet. Es ist $u_{\text{ind}2} = -\mathrm{d}\Phi_{21}/\mathrm{d}t$ mit Φ_{21} als jenem Teil des vom Strom i_1 in L_1 erzeugten magnetischen Flusses, der von dem zweiten Stromweg in L_2 umfaßt wird, so daß $\Phi_{21} = M_{21} \cdot i_1$.
Die Gegeninduktivität hängt von der geometrischen Form, den Abmessungen, der gegenseitigen Lage der beiden Stromkreise und der Permeabilität des Mediums ab. Entsprechend ist $\Phi_{12} = M_{12} \cdot i_2$ der Teil des von i_2 in L_2 erzeugten magnetischen Flusses, der von L_1 umfaßt wird.
Im idealen Übertrager, in dem sich beide Spulen auf einem gemeinsamen Kern befinden, werden beide Spulen vom gleichen magnetischen Fluß Φ durchsetzt. Die in ihnen induzierten Spannungen sind deshalb ihren Windungszahlen proportional: $u_1/u_2 = n_1/n_2$.
Da beim idealen Übertrager die gesamte Energie von der Primär- auf die Sekundärwicklung übertragen wird, müssen sich die entsprechenden Stromstärken zu den Windungszahlen reziprok verhalten: $i_2/i_1 = n_1/n_2$.
Netztransformatoren (für technischen Wechselstrom) und Übertrager (für nieder- und hochfrequente Ströme) sind die wichtigsten Anwendungen von Gegeninduktivitäten.

8.1.2. Elektronenröhren

Elektronenröhren gehören zu den vakuumelektronischen Bauelementen. In einem evakuierten Glaskolben wird ein Elektronenstrom von einer geheizten Katode emittiert, durch eine oder mehrere metallische Elektroden (Gitter) mittels veränderbarer Potentiale gesteuert und von einer gegenüber der Katode positiv vorgespannten Elektrode, der Anode aufgenommen. Um bei relativ niedrigen Katodentemperaturen hohe Emissionsstromdichten zu erreichen, sind niedrige Austrittsarbeiten des wirksamen Katodenmaterials nötig.

8.1.2.1. Dichtegesteuerte Elektronenröhren. Ist die Laufzeit der Elektronen zwischen Katode und Anode klein im Vergleich zur Schwingungsdauer des an der Steuerelektrode, z. B. einer Dreielektrodenröhre (Triode), anliegenden Signals, so wird durch die Signalspannung die zur Anode gelangende Stromdichte gesteuert. Durch die Halbleiterelektronik sind dichtegesteuerte Elektronenröhren, die die ersten steuerbaren elektronischen Bauelemente waren, weitgehend verdrängt worden und werden meist nur noch als Senderöhren hoher Leistung verwendet.

8.1.2.2. Laufzeitröhren. Wird die Periodendauer des Steuersignals so kurz, daß sie von der Größenordnung der Laufzeit der Elektronen zwischen Katode und Anode ist, versagt die Dichtesteuerung. In Laufzeitröhren bilden gerade diese Laufzeiteffekte die zugrundeliegenden Wirkprinzipien. Wird ein Elektronenstrahl konstanter Geschwindigkeit einem steuernden hochfrequenten elektrischen Feld ausgesetzt, so führt dies zu einer Steuerung der Elektronengeschwindigkeit. Je nach der Phase des Steuerfeldes beim Elektronenaustritt stellen sich periodisch über die Geschwindigkeitssteuerung Verdichtungen und Verdünnungen der Elektronenkonzentration ein (Bildung von „Elektronenpaketen").

Bei *Lauffeldröhren* (z. B. beim Magnetron) erfolgt die Wechselwirkung des Elektronenstroms mit der elektromagnetischen Schwingung über ein mitlaufendes elektrisches Feld, während sie bei *Triftröhren* (Klystrons) in einem räumlich eng begrenzten Elektrodensystem stattfindet. Laufzeitröhren dienen als Verstärker- und Senderöhren im Frequenzbereich bis zur Größenordnung von einigen 10^2 GHz.

8.1.2.3. Katodenstrahlröhren. Eine nicht geringere Bedeutung als die Laufzeitröhren hat eine dritte Gruppe der Elektronenröhren, die der Katodenstrahlröhren. Zu ihnen zählen Oszillografenröhren, Bildwiedergabe- und Bildaufnahmeröhren.

Oszillografenröhren sind Signalformwandler und erzeugen aus elektrischen Signalen optische (Abb. 8.1.2–1). Der

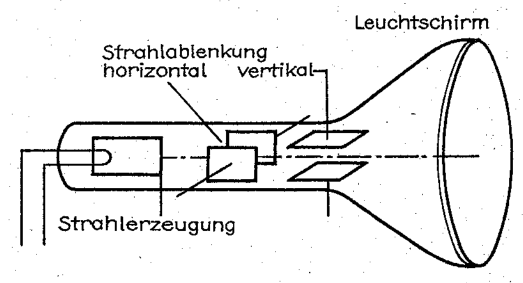


Abb. 8.1.2–1 Schematischer Aufbau einer Katodenstrahlröhre mit Strahlerzeugung, horizontaler und vertikaler Strahlablenkung und Leuchtschirm

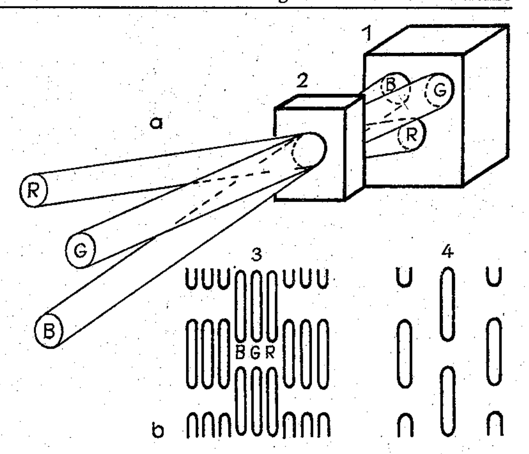


Abb. 8.1.2–2 Bildwiedergabe für Farbfernsehen; (a) Delta-Röhre: Anordnung von Leuchtschirm (1), Lochmaske (2) und Dreistrahlsystem, (b) Inline-Röhre: Anordnung der Leuchtstoffstreifen auf dem Leuchtschirm (3) und Schlitzmaske (4)

Elektronenstrahl verläßt eine indirekt geheizte Oxidkatode. Seine Intensität kann durch die Spannung zwischen einer Lochblende (Wehnelt-Zylinder) und der Katode gesteuert werden. Durch ein elektronenoptisches Linsensystem wird der Elektronenstrahl beschleunigt und zu einem Brennfleck auf dem Leuchtschirm der Röhre fokussiert. Zwei senkrecht aufeinanderstehende Plattenpaare lenken den Elektronenstrahl entsprechend der an ihnen liegenden Spannung ab. Legt man an das vertikale Plattenpaar eine periodische, zeitproportionale Spannung (Sägezahnform) und an das horizontale Plattenpaar die periodisch veränderliche Meßspannung, so zeigt der Leuchtschirm deren zeitlichen Verlauf. Katodenstrahloszillografen sind wichtige Hilfsmittel der Meßtechnik.

Bildwiedergaberöhren für Schwarzweißfernsehen arbeiten ähnlich wie Oszillografenröhren. Die Fokussierung des Elektronenstrahls kann sowohl elektrostatisch wie bei Oszillografenröhren wie auch durch konzentrisch um den Bildröhrenhals angeordnete Magnete erfolgen. Zur periodischen, zeilenweisen Führung des Elektronenstrahls über den Leuchtschirm dient ein Magnetfeld, das von zwei flachen Ablenkspulenpaaren erzeugt wird, die senkrecht zum Strahl und senkrecht zueinander angeordnet sind. Sie werden als Ablenkeinheit von außen über den Bildröhrenhals geschoben. Während der zeilenweisen Ablenkung erfolgt die Intensitätssteuerung des Elektronenstrahls entsprechend dem übertragenen Bildinhalt. Die sich ändernde Intensität des Elektronenstrahls bewirkt auf dem Bildschirm eine dazugehörige Leuchtdichte. Die Leuchtstoffe (Luminophore) für den Bildschirmbelag, die Zinksulfid und Cadmiumsulfid enthalten, sind so zusammengesetzt, daß ihr Nachleuchten (Phosphoreszenz) beim nächsten Überschreiben einer Zeile hinreichend abgeklungen ist. Um die Länge des Bildröhrenhalses in Grenzen zu halten, werden in der Praxis beträchtliche Strahlablenkwinkel (110° in der Bildschirmdiagonale) beherrscht.

Bildwiedergaberöhren für Farbfernsehen. Der Leuchtschirm solcher Röhren enthält in rasterartiger Verteilung Leuchtstoffe der Farben rot, grün und blau. Die drei Farben

des Leuchtschirms werden von drei getrennt steuerbaren Elektronenstrahlen angeregt. Ihre Fokussierung geschieht auf elektrostatischem Wege. Die Anordnung der Strahlsysteme kann auf zwei Arten erfolgen (Abb. 8.1.2–2). Entweder befinden sich die Strahlsysteme an den Ecken eines gleichseitigen Dreiecks (Deltaröhre) oder sie liegen in einer Linie (Inlineröhre). Damit der einer bestimmten Farbe zugeordnete Elektronenstrahl nur Rasterpunkte dieser Farbe auf dem Leuchtschirm anregt, ist in beiden Fällen eine Maske mit geeigneten Öffnungen vor dem Leuchtschirm anzubringen. Der Leuchtschirm der Deltaröhre, der aus drei Punktrastern mit ≈300 000 Leuchtpunkten je Grundfarbe besteht, erfordert eine Lochmaske mit der gleichen Anzahl von Löchern wie die Anzahl der Leuchtpunkte eines Rasters. Der Leuchtschirm der Inlineröhre hat anstelle der Punktraster drei Streifenraster der Grundfarben. Die Zuordnung der Strahlen zu den entsprechenden Luminophoren erfolgt mit Hilfe einer Schlitzmaske. Schlitzmaskenröhren bieten einige Vorteile bei Herstellung und Anwendung.

Bildaufnahmeröhren. In ihnen werden optische Bilder in elektrische Signale umgewandelt. Mit Hilfe des inneren fotoelektrischen Effekts erzeugt die optische Abbildung auf einer Speicherschicht aus z. B. Selen oder Antimontrisulfid ein Ladungsbild, das zeilenweise von einem scharf gebündelten Elektronenstrahl abgetastet wird. Der momentan und lokal auftretende Entladestrom kann über eine leitende Fläche abfließen und stellt das elektrische Signal der Helligkeit des entsprechenden Bildpunktes dar. Für die Herstellung von Bildaufnahmeröhren werden die Möglichkeiten mikroelektronischer Technologien genutzt.

Durch geeignete Halbleiterwerkstoffe kann man die Empfindlichkeit der Röhren im Infrarotbereich erhöhen, so daß auch Wärmestrahlungen dargestellt werden können.

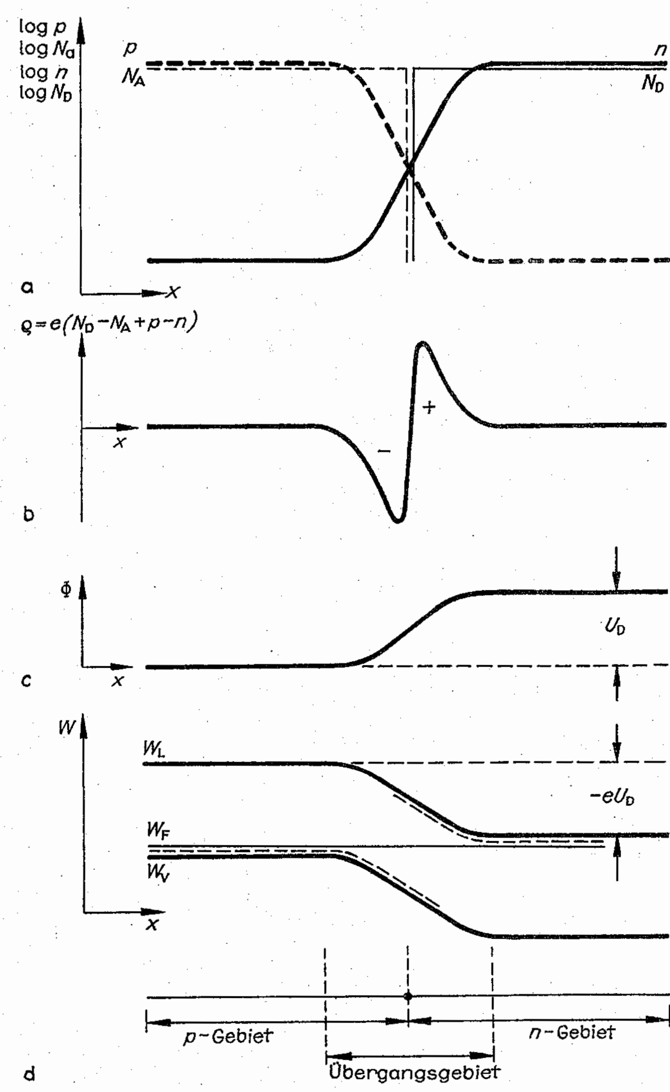


Abb. 8.1.3–1 *pn*-Übergang im stromlosen Zustand, (a) Ladungsträgerkonzentration, (b) Raumladung, (c) Diffusionsspannung, (d) Energiebändermodell

8.1.3. Halbleiterbauelemente

8.1.3.1. ***pn*****-Übergang.** Die Vielfältigkeit der elektronischen Anwendung von Halbleitern beruht nicht so sehr auf der Vielzahl der vorhandenen halbleitenden Substanzen, sondern in weit stärkerem Maße auf der umfassenden Nutzung von Eigenschaften einer Grundstruktur, der Kombination von p- und n-leitenden Gebieten desselben Halbleitermaterials, des pn-Überganges. Die in Bedeutung und Anwendungsbreite alle anderen technisch verwertbaren Halbleitermaterialien übertreffende Substanz ist der Elementhalbleiter Silicium. Der pn-Übergang, an dessen Grenzfläche p-leitendes und n-leitendes Material zusammenstoßen, stellt in Richtung und Gegenrichtung fließenden Strömen stark unterschiedliche Widerstände entgegen.

Das p-Gebiet wird durch bewußten Einbau (Dotierung) eines Elements der III. Hauptgruppe des Periodensystems der Elemente (z. B. Bor, Aluminium, Indium, Gallium) in den extrem reinen Elementhalbleiter Silicium hergestellt. Die Dotierungsatome erzeugen Störstellen innerhalb der verbotenen Zone in der Nähe der oberen Valenzbandkante. Diese Störstellen, die als Akzeptoren bezeichnet werden, nehmen Elektronen aus dem Valenzband auf, binden diese an die Dotierungsatome, die dadurch zu negativ geladenen Ionen werden, und lassen bewegliche Defektelektronen im Valenzband zurück, erzeugen also p-Leitung. Das n-Gebiet entsteht durch Dotierung mit einem Element der V. Hauptgruppe des Periodensystems (z. B. Phosphor, Arsen, Antimon), das als Donatoren bezeichnete Störstellen in der Nähe der unteren Kante des Leitungsbandes bildet. Jeder Donator gibt leicht ein Elektron an das Leitungsband ab. Das Donatoratom wird zu einem positiv geladenen Ion. Wenn alle Störstellen für p- bzw. n-Leitung in den entsprechenden Gebieten gesorgt haben und die Störstellen nicht mehr zu einer weiteren Erhöhung der Ladungsträgerkonzentration beitragen können, spricht man von Störstellenerschöpfung. Die Eigenleitung sei vernachlässigbar klein.

Am pn-Übergang soll die Akzeptorkonzentration N_A von ihrem im p-Gebiet konstanten Wert (von beispielsweise $N_A = 10^{16}\,\mathrm{cm}^{-3}$) abrupt auf 0 zurückgehen, ebenso die Donatorkonzentration im n-Gebiet (Abb. 8.1.3–1). Diesen plötzlichen Übergang der fest ins Kristallgitter eingebauten Störstellenatome werden aber keineswegs die von ihnen bereitgestellten beweglichen Ladungsträger im Gleichgewichtszustand einnehmen. Der steile Konzentrationsgradient der beweglichen Ladungsträger am Übergang hat einen Diffusionsstrom in das Gebiet zur Folge, in dem sie jeweils Minoritätsträger sind, d. h. es gibt einen Diffusionsstrom der im n-Gebiet in der Nähe des Übergangs befindlichen Elektronen in das p-Gebiet und der im p-Gebiet in der Nähe des Übergangs befindlichen Defektelektronen in das n-Gebiet. Diese Diffusionsströme sind, dem 1. Fickschen Diffusionsgesetz entsprechend, dem Konzentrationsgradienten direkt proportional. Die Ladungsträgerdiffusion hat im Übergangsbereich eine Raumladung zur Folge. Diese Raumladung ist im p-Gebiet negativ. Sie rührt von den fest im Kristallgitter eingebauten, negativ geladenen Akzeptorionen her (Abb. 8.1.3–1b). Im n-Gebiet bleibt von den positiven Donatorionen eine positive Raumladung zurück. Zwischen den ins Gitter eingebauten Ionen unterschiedlichen Vorzeichens hat sich ein mit dem pn-Übergang fest verbundenes elektrisches Feld aufgebaut. Dieses elektrische Feld wirkt auf die beweglichen Ladungsträger und erzeugt dadurch einen Feldstrom, der dem Diffusionsstrom entgegengerichtet ist. Im Gleichgewichtsfall sind beide Ströme für jede Ladungsträgersorte, also Elektronen und Defektelektronen gleich groß und heben sich gegenseitig auf. Die durch die innere Feldstärke über das Raumladungsgebiet hinweg erzeugte Potentialschwelle wird als Diffusionsspannung U_D bezeichnet. Die Diffusionsspannung U_D hängt von der Störstellenkonzentration ab. Es gilt $U_D = (kT/e) \ln N_A N_D / n_i^2$ (n_i = Inversionsdichte; im stromlosen Zustand des pn-Übergangs ist $n_i^2 = np$).

In der Darstellung des Energiebändermodells für den pn-Übergang liegt die Fermi-Energie, bei deren Höhe die Besetzungswahrscheinlichkeit für einen Zustand 1/2 ist, im thermischen Gleichgewicht im p- und im n-Gebiet auf dem gleichen Wert. Die durch die Diffusionsspannung U_D gebildete Potentialschwelle muß somit zu einer Verbiegung der

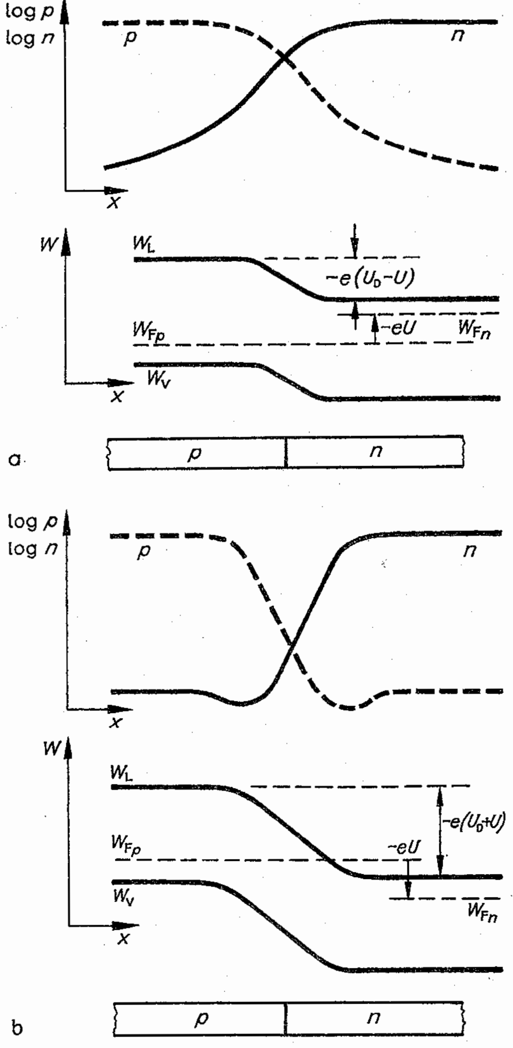


Abb. 8.1.3–2 Stromdurchflossener pn-Übergang, Ladungsträgerkonzentrationen und Bändermodell, (a) Flußpolung, (b) Sperrpolung

Energiebänder um $W = -eU_D$ führen. Die Bandverbiegung deutet auf das im Bereich der Raumladungszone liegende innere elektrische Feld.
Soll z. B. ein Elektron, das die Energie der unteren Leitungsbandkante W_L aufweist, vom n-Gebiet in das p-Gebiet gelangen, so muß ihm gerade die potentielle Energie zugeführt werden, um die die Bandkanten gegeneinander verschoben sind (vgl. Abb. 8.1.3–1).
Wird an den pn-Übergang eine äußere Spannung angelegt, so ist deren Wirkung von ihrer Polarität abhängig. Legt man den Pluspol der Spannungsquelle an das n-Gebiet und den Minuspol an das p-Gebiet, dann wird das durch die Raumladung bedingte innere elektrische Feld, für das die Diffusionsspannung kennzeichnend ist, vergrößert. Das heißt gleichzeitig, daß das Übergangsgebiet noch stärker von beweglichen Ladungsträgern entblößt wird und, daß sich die Bandverbiegung vergrößert. Bei dieser Polung des äußeren Feldes entsteht der in Sperrichtung vorgespannte pn-Übergang. Das trifft für die Majoritätsladungsträger, also die Elektronen im n- und die Löcher im p-Gebiet zu, die unter dieser Bedingung keinen Strom liefern können. Die Minoritätsladungsträger, also die Elektronen im p-Gebiet und die Löcher im n-Gebiet, die in gewisser Konzentration, z.B. auf thermischem Wege erzeugt werden können, bilden den sehr kleinen temperaturabhängigen Sättigungssperrstrom.
Verbindet man dagegen den Pluspol der Spannungsquelle mit dem p-Gebiet und den Minuspol mit dem n-Gebiet, so trägt diese Polung zur Schwächung des inneren Feldes bei. Die durch die Diffusionsspannung U_D bedingte Potentialschwelle wird, ebenso wie die Bandverbiegung, verringert.
Da im stromlosen Zustand Diffusionsstrom und Feldstrom einander kompensierten, ist dieses Gleichgewicht jetzt nicht mehr aufrechtzuerhalten. Das Raumladungsgebiet wird mit beweglichen Ladungsträgern überschwemmt und es können ständig neue Ladungsträger nachgeliefert werden. Von beiden Seiten aus erfolgt eine Injektion von Minoritätsladungsträgern über die Gleichgewichtswerte (Abb. 8.1.3–2a) hinaus. Mit größer werdendem Abstand vom Übergangsgebiet reduzieren sich jedoch die Strombeiträge der injizierten Minoritätsträger. Der Gesamtstrom wird außerhalb des Übergangsgebietes überwiegend von Majoritätsladungsträgern gebildet. Die Strom-Spannungs-Kennlinie eines pn-Übergangs ist angenähert darstellbar durch $I = I_{sp}\{\exp(eU/kT) - 1\}$ (Abb. 8.1.3–3). Die Spannung kT/e beträgt bei Raumtemperatur ≈ 25 mV. Der Sättigungssperrstrom liegt für kleinflächige Silicium-pn-Übergänge in der Größenordnung von Nanoampere, für Germanium bei $\approx 10\,\mu A$.
Der pn-Übergang hat eine Kapazität C. Bei Sperrpolung und zunehmender Ladungsträgerextraktion aus dem Übergangsgebiet verringert sich C. Die indirekte Proportionalität zwischen Sperrspannung U und der Sperrschichtkapazität C werden bei der *Kapazitäts-* oder *Varaktordiode* (vgl. 8.1.3.2.) genutzt, die bis zu Frequenzen von mehr als 100 GHz betrieben werden kann.
Bei sehr hohen Sperrspannungen setzt durch nichtlineare Prozesse ein Stromdurchbruch ein, der ebenfalls in Bauelementen (z. B. der *Z-Diode*, vgl. 8.1.3.2.) technisch angewandt wird.

8.1.3.2. Halbleiterdioden sind Bauelemente mit zwei Anschlüssen und nichtlinearer Strom-Spannungs-Kennlinie. Wird an eine Reihenschaltung von Diode und Widerstand eine Spannung in Flußrichtung angelegt, so zeigt die Diode einen sehr kleinen Durchlaßwiderstand. Bei Umkehrung der Polarität der Spannung ist der Widerstand hoch, der vernachlässigbar kleine Sperrstrom wächst mit der Temperatur. Diese Gleichrichtereigenschaften von Halbleiterdioden werden durch das Verhalten von pn-Übergängen bestimmt. Halbleiterdioden stellen im wesentlichen pn-Übergänge dar. Eine Ausnahme bilden Schottky-Dioden, deren Gleichrichtereigenschaften durch einen Metall-Halbleiter-Kontakt hervorgerufen werden. Nach ihrer Herstellungsform ist zwischen Spitzen- und Flächendioden zu unterscheiden (Abb. 8.1.3–4). *Spitzendioden* bestehen z. B. aus einem n-leitendem Plättchen, auf das eine Metallspitze aufgesetzt wurde. Mit einem kurzen Stromstoß hoher Stärke werden im Halbleiter durch Formierung Akzeptoren erzeugt. Die Spitzendiode ist ein Nachfahre des PbS-Detektors aus der Zeit beginnender Rundfunkübertragungen.
Flächendioden lassen sich durch Einlegieren der Störstellenleitung erzeugenden Substanz, durch Eindiffusion oder durch Ionenimplantation herstellen. Verbreitet ist die Silicium-Epitaxie-Planar-Technologie.

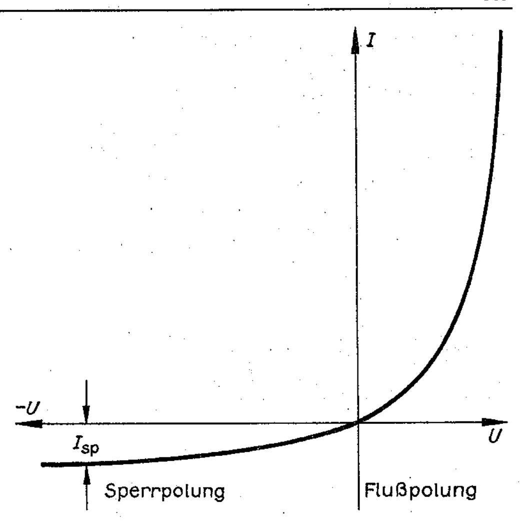


Abb. 8.1.3–3 Strom-Spannungs-Kennlinie eines pn-Überganges

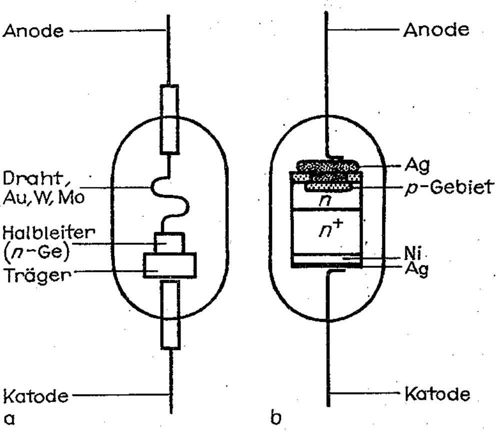
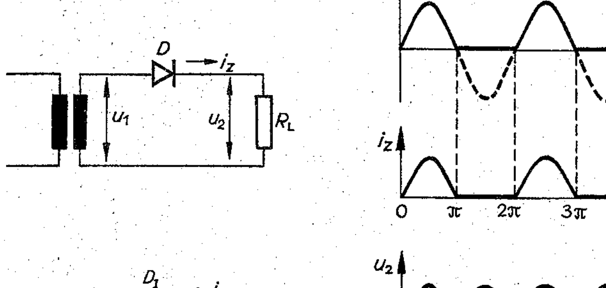
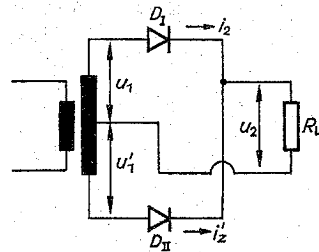


Abb. 8.1.3–4 Diodenaufbau (schematisch), (a) Spitzendiode, (b) Flächendiode

Dioden haben vielfältige Einsatzgebiete. Am wesentlichsten ist die Verwendung als Gleichrichter für Ströme in Stromversorgungseinrichtungen und zur Gleichrichtung hochfrequenter Ströme sowie als Schalter in der Digitalelektronik. Die sich um Größenordnungen unterscheidenden Stromstärken in Fluß- und Sperrichtung und das kurzfristige Reagieren auf Änderungen der Steuergröße liegen zum Verbraucher parallel geschalteten Ladekondensator geglättet werden. Eine angeschlossene Siebkette reduziert den Wechselspannungsanteil weiter.

Eine Vergrößerung der Gleichspannung ergibt sich durch Zweiweggleichrichter. Sie ist mit einer Gegentaktschaltung zu verwirklichen, für die an der Sekundärwicklung eine Mittelanzapfung vorhanden sein muß (Abb. 8.1.3–5b).





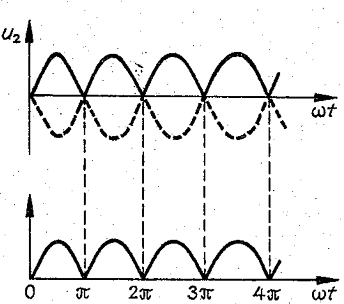


Abb. 8.1.3–5 Gleichrichtung von Wechselströmen, (a) Einweggleichrichtung, (b) Zweiweggleichrichtung

der Nutzung als schnelle Schalter in Datenverarbeitung und Steuer- und Regelungstechnik zugrunde. Als Beispiele für Sonderanwendungen seien die Z-Diode, die Kapazitätsdiode, die Tunneldiode, Foto- (vgl. 8.1.6.2.), Lumineszenz- (vgl. 8.1.6.4.) und Laserdiode (vgl. 8.1.6.5.) genannt.

Richt- und Schaltdioden. Die Mehrzahl der diskret oder in integrierten Strukturen hergestellten Dioden werden zur Gleichrichtung von Strömen bzw. als Schalter benutzt. Die einfachste Schaltung zur Gleichrichtung technischer Wechselströme ist die Einweggleichrichtung (Abb. 8.1.3–5a). Während der positiven Halbwelle der Wechselspannung $u = \hat{U}\sin\omega t$ befindet sich der Gleichrichter D in Flußpolung und wegen des geringen Widerstands des Gleichrichters liegt fast die gesamte Spannung am Verbraucher R_L. Während der negativen Halbwelle der Spannung u hat der Gleichrichter Sperrpolung, in der wegen des hohen Sperrwiderstands praktisch kein Strom fließt, also am Verbraucher R_L auch keine Spannung abfallen kann. Die negative Halbwelle der Spannung liegt als Sperrspannung am Gleichrichter.

Die über die Zeitdauer einer Periode T gemittelte Spannung bei Einweggleichrichtung ${}_1\bar{U}$ ist

$${}_1\bar{U} = 1/T\int_0^{T/2} u(t)\,\mathrm{d}t = \hat{U}\omega/(2\pi)\int_0^{\pi/\omega} \sin\omega t\,\mathrm{d}t = (1/\pi)\hat{U} = 0{,}318\,\hat{U}.$$

Vergleicht man ${}_1\bar{U}$ mit dem quadratischen Mittelwert der Spannung, der effektiven Spannung

$$U_{\text{eff}} = \sqrt{(1/T)\int_0^T U^2(t)\mathrm{d}t} = \hat{U}/\sqrt{2},$$

so wird ${}_1\bar{U} = 0{,}318\sqrt{2}\;U_{\text{eff}} = 0{,}45\,U_{\text{eff}}$.

Die erzielte pulsierende Gleichspannung kann durch einen

Der Lastwiderstand R_L liegt zwischen der Mittelanzapfung der Sekundärwicklung und der Verbindung beider Katoden. Da die Spannungen u_1 und u'_1 auf die Mittelanzapfung bezogen um π in der Phase verschoben sind, ist im Verlaufe der jeweils ersten Halbwelle D_I leitend und D_{II} gesperrt und im Verlaufe der jeweils zweiten Halbwelle einer Periode ist D_I gesperrt und D_{II} leitend. Die mittlere Spannung bei Zweiweggleichrichtung ${}_2\bar{U}$ ist entsprechend

$${}_2\bar{U} = 2/T\int_0^{T/2} u(t)\,\mathrm{d}t = 2\hat{U}/\pi = 0{,}637\,\hat{U} = 0{,}90\,U_{\text{eff}}.$$

Die Rückgewinnung eines Signals, das einer hochfrequenten Trägerschwingung aufgeprägt wurde, bezeichnet man als Hochfrequenzgleichrichtung oder Demodulation. Diese Aufgabe wird am häufigsten ebenfalls mit Richtdioden durchgeführt. Eine Diodendemodulationsschaltung für eine amplitudenmodulierte Trägerfrequenz zeigt Abb. 8.1.3–6. Das wiederzugewinnende Signal ist mit der Hüllkurve identisch, die die sich ändernden Amplituden der Trägerschwingung umgibt. Der Arbeitspunkt der Diode kann auf $U_D = 0$ eingestellt werden, liegt jedoch meist im Sperrbereich. Damit werden die negativen Hälften der amplitudenmodulierten Schwingungen abgeschnitten. An R_L liegt die pulsierende Gleichspannung an, die den Kondensator C auflädt. Die Zeitkonstante $\tau = R_L C$ muß so gewählt werden, daß die Spannung R_L den zeitlichen Verlauf des Signals gut wiedergibt, der Trägerfrequenz jedoch nicht mehr folgen kann. Unter dieser Bedingung ergibt sich an R_L der Spannungsverlauf der Hüllkurve. Da die positiven Hälften der Trägerschwingungen eine Gleichspannungskomponente enthalten, sorgt der Kondensator für deren Abtrennung. Für zu kleine HF-Amplituden unterscheidet sich die Diodenkenn-

linie in ihrem Verlauf am Nullpunkt nicht deutlich zwischen Fluß- und Sperrzustand. Deswegen ist eine ausreichend große Amplitude der Hochfrequenzschwingung für die Demodulation erforderlich.
Schaltdioden werden vorzugsweise in der Digitaltechnik eingesetzt. Sie haben die Funktion elektronischer Schalter auszuüben. Es interessiert besonders das Verhalten in den beiden Schaltzuständen „Schalter geschlossen" und „Schalter offen" sowie die Eigenschaften beim Übergang von einem Schaltzustand in den anderen. Als Eigenschaften eines idealen Schalters, die anzustreben sind, wird folgendes Verhalten angesehen: Im Zustand „Schalter geschlossen" hat der Schalter den Widerstand null, d. h. an ihm tritt kein Spannungsabfall auf. Im Zustand „Schalter offen" ist der über den Schalter fließende Strom null. Der Übergang von einem Schaltzustand in den anderen soll nach Auslösung der Schaltoperation ohne zeitliche Verzögerung erfolgen. Diesen Forderungen werden die Schaltdioden angenähert gerecht. Zudem werden in der Digitaltechnik den beiden Signalzuständen „Schalter offen" oder „Schalter geschlossen" bzw. „hohes Potential" oder „niedriges Potential" keine diskreten Werte, sondern Intervalle zugeordnet. Jedoch sinkt im Zustand „Schalter geschlossen" der Widerstand der in Flußrichtung gepolten Diode nicht auf null und im Zustand „Schalter geöffnet" fließt der Diodensperrstrom. Auch bis zur Einstellung des neuen Schaltzustandes vergeht eine, wenn auch sehr kurze, Zeitdauer. Von beiden Übergängen nimmt der Übergang von Fluß- zum Sperrfall die längere Dauer in Anspruch, da die im Flußfall unwirksamen Minoritätsträger nach der Sperrpolung erst durch Rekombination unwirksam gemacht werden müssen. Unter Ausnutzung der Abhängigkeit des Diodenverhaltens von der Polarität der angelegten Spannung lassen sich Diodenschaltungen aufbauen, mit denen logische Verknüpfungen binärer Größen durchgeführt werden können.
Für die Verknüpfung zweier binärer Eingangsgrößen durch die logische Operation UND ist die Schaltung in Abb. 8.1.3–7a geeignet. Befindet sich z. B. der Eingang E_1 auf Erdpotential, dann fließt ein Strom über die in Flußpolung befindliche Diode D, der Durchlaßwiderstand der Diode ist klein gegen den Widerstand R und am Ausgang A ist Erdpotential. Das gilt auch, wenn der andere Eingang E_2 oder beide Eingänge auf Erdpotential liegen. Erst wenn E_1 und E_2 auf positivem Potential liegen, sind beide Dioden gesperrt und A liegt auf positivem Potential, wie das für die UND-Verknüpfung gefordert ist. Für die ODER-Verknüp-

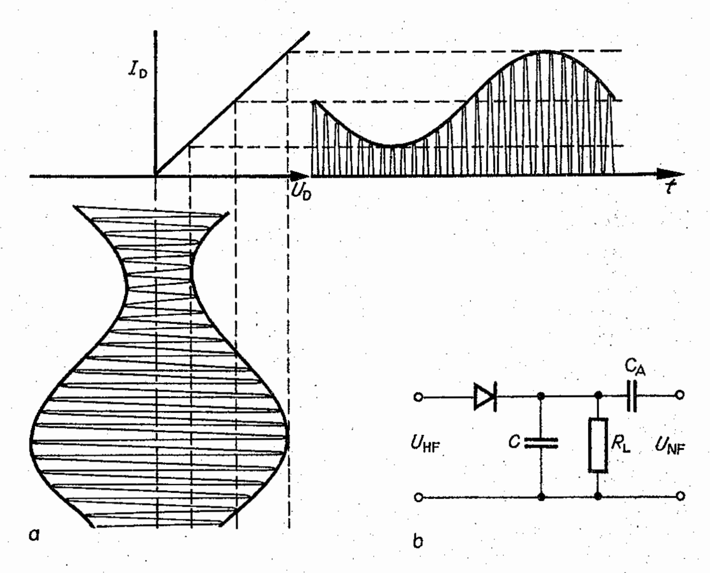


Abb. 8.1.3–6 Demodulation, (a) Demodulation amplitudenmodulierter Schwingungen an einem Bauelement mit Kennlinienknick, (b) Diodendemodulationsschaltung

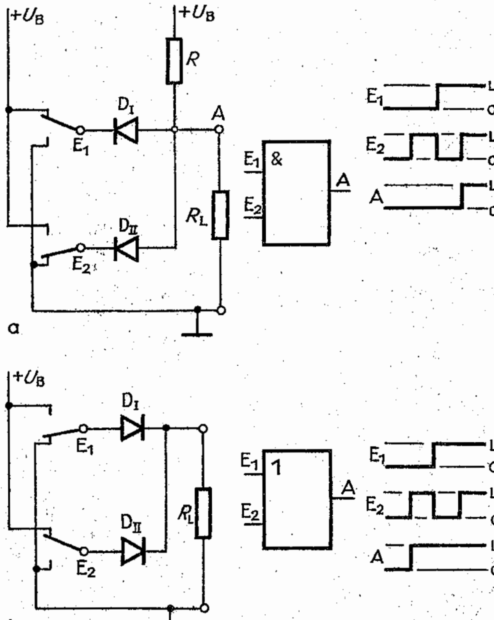


Abb. 8.1.3–7 Logische Verknüpfung zweier binärer Größen in Diodenschaltungen mit Schaltskizze, Verknüpfungssymbol und Signalverläufen, (a) UND-Schaltung (Konjunktion), (b) ODER-Schaltung (Disjunktion)

fung (Abb. 8.1.3–7b) muß nur ein Eingang auf positivem Potential liegen, damit auch der Ausgang diesen Zustand annimmt. Nur wenn beide Eingänge auf Erdpotential liegen, liegt auch der Ausgang auf Erdpotential. Die für logische Operationen unerläßliche Negation ist mit Dioden nicht zu realisieren. Das wird erst durch Hinzunahme eines Transistors als Negator möglich. Es entsteht das System der Dioden-Transistor-Logik (DTL). Größere Verbreitung hat das System der Transistor-Transitor-Logik (TTL) gefunden (vgl. 8.2.4.2.). Schaltdioden werden auch zu Tor- und Begrenzerschaltungen eingesetzt, bei denen bestimmten Spannungsverhältnissen Schaltvorgänge zugeordnet sind.

Z-Diode. Je nach beabsichtigtem Anwendungszweck lassen sich *pn*-Übergänge mit typischen Eigenschaften herstellen. Zu ihnen zählen die Z-Dioden, bei denen von einer bestimmten Sperrspannung an ein steil ansteigender Strom einsetzt. Für diese als Spannungsdurchbruch bezeichnete Erscheinung gibt es zwei Ursachen. Die erste Ursache ist eine innere Feldemission aufgrund einer hohen inneren Feldstärke (einige 10^5 V/cm) infolge eines hochdotierten, schmalen *pn*-Übergangs. Dadurch werden gebundene Elektronen von ihren Atomen getrennt. Die Elektronen gelangen unter Beibehaltung ihrer Energie mit Hilfe des Tunneleffektes (vgl. 8.1.3.2.) in das Leitungsband. Die Voraussetzungen sind die gleichen wie bei der Tunneldiode. Der Spannungsdurchbruch infolge innerer Feldemission (Zener-Durchbruch) tritt für Silicium bei vergleichsweise kleinen Sperrschichtbreiten und bereits bei kleinen Sperrspannungen zwischen $0{,}5\,\mathrm{V} \leqq U_{sp} \leqq 4\,\mathrm{V}$ auf. Die zweite Ursache für den Spannungsdurchbruch ist Stoßionisation. Die Ladungsträger werden im elektrischen Feld zwischen zwei Zusammenstößen mit dem Kristallgitter derart beschleunigt, daß ihre Energie ausreicht, Valenzelektronen aus ihren Bindungen zu reißen, die ihrerseits beschleunigt werden und zu Stoßionisation in der Lage sind. Dadurch entsteht eine kettenreaktionsartige Ladungsträgervervielfachung, die als Lawinendurchbruch bezeichnet wird. Der Spannungsdurchbruch infolge Stoßionisation vollzieht sich bei größeren Sperrschichtbreiten und bei höheren Spannungen (für Silicium: $5\,\mathrm{V} \leqq U_{sp} \leqq 200\,\mathrm{V}$) als der Zener-Durchbruch. Der Einsatz des Lawinendurchbruches ist schärfer als der des

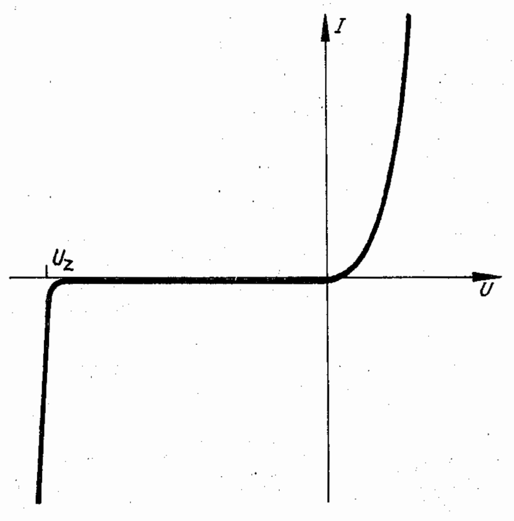


Abb. 8.1.3–8 Strom-Spannungs-Kennlinie einer Z-Diode

Zener-Durchbruches, was am schärferen Knick der Strom-Spannungs-Kennlinie (Abb. 8.1.3–8) deutlich wird.

Der Temperaturkoeffizient der Durchbruchsspannung hat für beide Mechanismen unterschiedliches Vorzeichen. Für kleine Durchbruchsspannungen (Zener-Durchbruch) ist der Temperaturkoeffizient negativ, für größere Durchbruchsspannungen (Lawinendurchbruch) positiv.

Kennzeichnend für Z-Dioden ist die Durchbruchsspannung U_Z, die mit der Herstellungstechnologie zwischen $\approx 1\,\mathrm{V} \leqq U_Z \leqq 200\,\mathrm{V}$ eingestellt werden kann, deren Temperaturkoeffizient und der differentielle Widerstand $r_Z = \mathrm{d}U_Z/\mathrm{d}I_Z$ im Durchbruchsbereich. r_Z ist klein und liegt zwischen wenigen Ohm und einigen Dutzend Ohm. Diesem Umstand verdankt die Z-Diode ihre Anwendung zur Spannungsstabilisierung und Spannungsbegrenzung, da Stromänderungen im Durchbruchsbereich entsprechend der Kennlinie nur von minimalen Spannungsänderungen begleitet sind.

Kapazitätsdiode (Varaktordiode). Der in Sperrichtung gepolte *pn*-Übergang bildet eine von beweglichen Ladungsträgern entblößte Zone aus, deren Breite sich mit wachsender Sperrspannung vergrößert (vgl. 8.1.3.1.). Diese Sperrschicht kann als das Dielektrikum eines Kondensators, der Sperrschichtkapazität des *pn*-Übergangs, betrachtet werden. Die Sperrschichtkapazität hängt gemäß 8.1.1.2. von der Fläche des *pn*-Übergangs und von der Breite der Sperrschicht ab. Da die Kapazität eines Plattenkondensators dem Plattenabstand indirekt proportional ist und der Plattenabstand der Sperrschichtbreite entspricht, folgt eine indirekte Proportionalität zwischen der Größe der Sperrspannung U und der Sperrschichtkapazität $C_S(U) = C_{S_0}(1 + U/U_D)^{-1/2+m}$ (U_D = Diffusionsspannung des *pn*-Übergangs, m = ein den Verlauf des Dotierungsprofils beschreibender Parameter, der z. B. für einen abrupten Übergang gleich null und für sich linear ändernde Dotierungskonzentrationen gleich 1 ist).

Obwohl im Grunde alle in Sperrichtung betriebenen Dioden als Varaktordioden verwendet werden können, werden bevorzugt Silicium-*pn*-Übergänge mit niedrigem Sperrstrom, hoher periodischer Spitzensperrspannung und speziell eingestelltem Störstellenprofil benutzt. Die Sperrschichtkapazitäten sind klein und liegen zwischen 1 pF und 100 pF. Je nach dem Dotierungsverlauf am *pn*-Übergang gelingen innerhalb der Variationsmöglichkeit der Sperrspannung Kapazitätsänderungen von 1:2 bis ≈1:30. Wegen der kleinen veränderlichen Kapazitäten eignen sich die Varaktordioden besonders für den Einsatz bei höheren Frequenzen zur Abstimmung von Schwingkreisen, insbesondere zur automatischen Frequenznachstimmung (AFC) in Rundfunkempfängern. Infolge der Nichtlinearität der Sperrschichtkapazität-Spannungs-Kennlinie eignet sich die Varaktordiode auch zur Frequenzvervielfachung. Eine parallel zu einem Schwingkreis liegende Varaktordiode sorgt für die Anregung harmonischer Schwingungen mit ganzzahligen Vielfachen der Grundschwingung.

Tunneldiode. Der Tunneleffekt der Quantenmechanik (vgl. 4.4.1.), der u. a. bei der kalten Emission von Elektronen aus Metalloberflächen unter dem Einfluß hoher elektrischer Feldstärken (Feldemission) und dem radioaktiven Zerfall von Atomkernen schwerer Elemente in Atomkerne des Heliums (α-Teilchen) und die verbleibenden Atomkerne (α-Zerfall) auftritt, kann unter speziellen Voraussetzungen auch an einem *pn*-Übergang als innere Feldemission herbeigeführt werden (Esaki 1958). Für diesen Zweck sind die Dotierungskonzentrationen im *p*- und *n*-Gebiet um etwa drei Größenordnungen höher als in normalen Halbleiter-

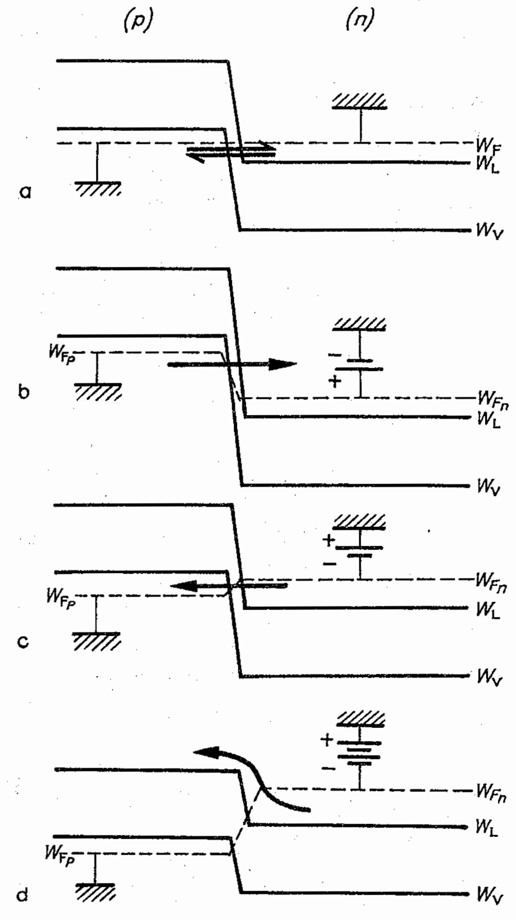


Abb. 8.1.3–9 Energiebandmodell einer Tunneldiode

dioden, also auf $\approx 10^{20}$ Fremdatome je cm^3, einzustellen. Dadurch entarten die Halbleitergebiete, und die Fermi-Grenze, bei $T = 0$ die Grenzenergie für besetzte und unbesetzte Zustände, liegt im stromlosen Fall im p-Gebiet unterhalb des oberen Valenzbandrands und im n-Gebiet oberhalb des unteren Leitungsbandrandes (Abb. 8.1.3–9a), also innerhalb der Energiebänder.
Die hohen Dotierungskonzentrationen sorgen dafür, daß die sich am pn-Übergang aufbauende Diffusionsspannung U_D (vgl. 8.1.3.2.) größer wird als die sich aus der Breite der verbotenen Zone W_L ergebende Spannung W_L/e.
Weiterhin ist ein hinreichend scharfer Übergang zwischen p- und n-dotiertem Gebiet zu gewährleisten. Durch Ladungsträgerdiffusion entsteht im Übergangsgebiet die an beweglichen Ladungsträgern verarmte Sperrschicht. Die hohen Dotierungskonzentrationen führen zu großen Raumladungen und die sehr geringe Dicke der entstehenden Sperrschicht von $\approx 10^{-2}$ µm bewirkt hohe elektrische Feldstärken bis zu 10^5 V/cm.
Das Verhalten der Tunneldiode bei unterschiedlicher Vorspannung wird anhand des Energiebändermodells (Abb. 8.1.3–9) deutlich. Soll der Tunneleffekt einsetzen, müssen auf der Seite der Potentialbarriere, auf der die Ladungsträger starten, besetzte Zustände vorhanden sein und auf der Seite jenseits der Barriere müssen freie Zustände existieren. Eine angemessene Durchtrittswahrscheinlichkeit ist bei geringer Barrierenbreite und hoher anliegender Feldstärke gegeben. Im Falle fehlender äußerer Spannung am pn-Übergang ist der resultierende Strom null (Abb. 8.1.3–9a), die Zahl der Elektronen des Valenzbandes aus dem p-Gebiet, die durch die Sperrschicht auf unbesetzte Zustände des Leitungsbandes des n-Gebiets tunnelt, ist gleich der Anzahl der Elektronen aus dem Leitungsband des n-Gebiets, die mit Hilfe des Tunneleffektes ohne Energieverlust auf freie Zustände des Valenzbandes des p-Gebiets gelangt.
Die vom p-Gebiet ins n-Gebiet kommenden Elektronen werden als Zener-Strom bezeichnet, bei entgegengesetzter Richtung spricht man vom Esaki-Strom.
Legt man an das n-Gebiet ein im Vergleich zum p-Gebiet positives Potential (Sperrpolung bei einer normalen Diode), dann erniedrigt sich die Energie der Ladungsträger des n-Gebiets gegenüber dem p-Gebiet (Vergrößerung der Bandüberlappung gegenüber Abb. 8.1.3–9a) und Elektronen aus dem p-Gebiet durchtunneln die Sperrschicht vorrangig in Richtung des n-Gebiets (Abb. 8.1.3–9b). Diese Tendenz liest man aus dem Bändermodell ab: Den besetzten Elektronenzuständen im p-Gebiet stehen im n-Gebiet unbesetzte Elektronenzustände gleicher Energie gegenüber.
Ein geringes negatives Potential am n-Gebiet (Durchlaßpolung bei einer normalen Diode) führt zu einer Verringerung der Bandüberlappung gegenüber dem stromlosen Fall, indem die Gesamtenergie des n-Gebiets angehoben wird (Abb. 8.1.3–9c). Nun stehen auf gleichem Energieniveau besetzten Elektronenzuständen im n-Gebiet unbesetzte im p-Gebiet gegenüber. Daraus resultiert ein Überwiegen des Elektronenstroms vom n- in das p-Gebiet.
Bei weiterer energetischer Anhebung des n-Gebietes durch Vergrößerung der Potentialdifferenz gegenüber Abb. 8.1.3–9c stehen jetzt den mit Elektronen besetzten Energiezuständen im n-Gebiet vorwiegend Energien im p-Gebiet gegenüber, die in der verbotenen Zone liegen, also nicht besetzt werden können (Abb. 8.1.3–9d). Gegenüber Abb. 8.1.3–9c muß eine Verringerung des Stromes eintreten. Eine weitere Vergrößerung der Potentialdifferenz im angegebenen Sinne mündet nun in das normale Diodenverhalten ein, die Potentialbarriere erniedrigt sich derart, daß Elektronenübergänge über die Barriere möglich werden, durch die der Diffusionsstrom normaler Dioden gekennzeichnet ist. Insgesamt ergibt sich für die Tunneldiode die in Abb. 8.1.3–10 dargestellte Strom-Spannungs-Kennlinie. I_H und I_T heißen Höcker- bzw. Talstrom und der Kennlinienteil 1 wird vorwiegend durch den Zener-Strom (Abb. 8.1.3–9b) und Teil 2 durch den Esaki-Strom (Abb. 8.1.3–9c) gekennzeichnet. Teil 3 der Kennlinie stellt den einsetzenden Diffusionsstrom in Flußrichtung (Abb. 8.1.3–9d) dar.
Entsprechend der Bandstruktur verwendbarer Halbleitermaterialien liegen U_H und U_T bei Werten, die kleiner sind als die der Breite der verbotenen Zone zuzuordnende Spannung. Höckerspannungen liegen bei 50 mV und Talspannungen zwischen 300 und 500 mV. Das Verhältnis von Höckerstrom I_H zu Talstrom I_T ist bei Tunneldioden aus Galliumarsenid (GaAs) am günstigsten und liegt maximal bei 60, für Germanium wird ≈ 10 und für Silicium 6 erreicht. Die Höckerströme können zwischen <1 mA und ei-

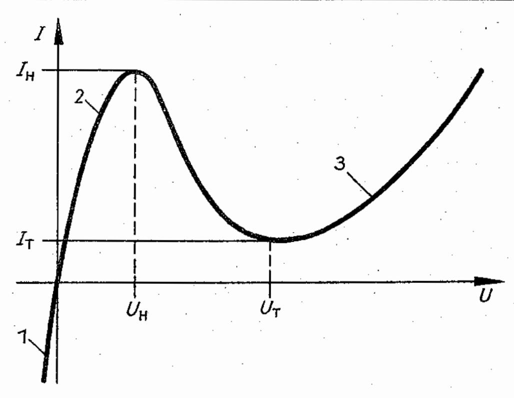


Abb. 8.1.3–10 Strom-Spannungs-Kennlinie einer Tunneldiode

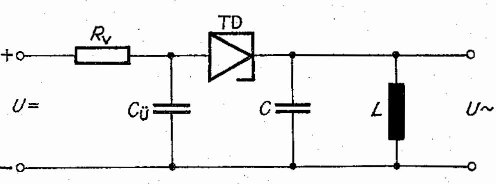


Abb. 8.1.3–11 Oszillator mit Tunneldiode TD

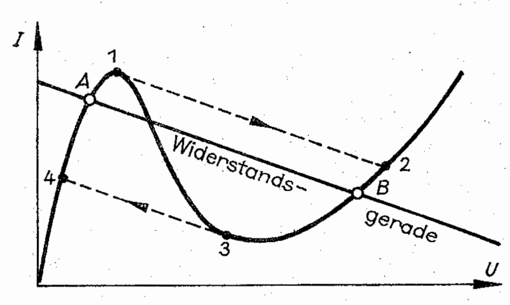


Abb. 8.1.3–12 Tunneldiode im Schalterbetrieb

nigen Ampere gewählt werden. Eine Abwandlung der Tunneldiode, bei der der Tunneleffekt ausschließlich in der „Sperrpolung“ auftritt, bei der also der Höcker fehlt, zeigt für kleine Spannungen demzufolge eine Umkehrung von Fluß- und Sperrichtung gegenüber herkömmlichen Halbleiterdioden. Sie trägt die Bezeichnung *Rückwärtsdiode* und wird zur Gleichrichtung von Wechselspannungen kleiner Amplituden verwendet. Das bemerkenswerteste Kennzeichen der Tunneldiode ist die Abnahme der Stromstärke bei wachsender Spannung zwischen Höcker und Tal, was einem negativen differentiellen Widerstand entspricht, der zwischen -10 und $-1\,000\,\Omega$ liegen kann. Sie eignet sich damit zur Reduzierung des Ohmschen Widerstands, kann also zur Entdämpfung von Schwingkreisen eingesetzt werden. Der Stromtransport unterhalb der Talspannung erfolgt wegen des Tunneleffektes nahezu verzögerungsfrei. Diese Eigenschaft wird zur Anwendung der Tunneldiode in Schaltungen für die Erzeugung hochfrequenter Schwingungen bis über 100 GHz ausgenutzt.

Eine äußere Induktivität und die Sperrschichtkapazität der Tunneldiode stellen die frequenzbestimmenden Größen dar. Infolge der geringen Dicke der Raumladungszone der Sperrschicht bilden sich Kapazitäten von einigen Mikrofarad je Quadratzentimeter, die man durch kleinflächige *pn*-Übergänge gering zu halten versucht. Abb. 8.1.3–11 zeigt eine Oszillatorschaltung mit einer Tunneldiode. Die Anordnung schwingt, wenn der Resonanzwiderstand des Schwingkreises größer ist, als der Betrag des negativen differentiellen Widerstandes der Tunneldiode. Die Schwingungsamplitude entspricht etwa der Talspannung. R_V stabilisiert den Arbeitspunkt der Diode im fallenden Kennlinienbereich und legt den im Mittel fließenden Strom fest. Die Gleichspannungsquelle ist für die Betriebsfrequenz mit der Kapazität $C_ü$ überbrückt. Oszillatoren mit Tunneldioden setzen nur sehr geringe Leistungen um, kommen allerdings auch mit minimalem Volumenbedarf aus.

Der beim Tunneleffekt sehr schnell vor sich gehende Ladungsträgerübergang gestattet den Einsatz der Tunneldiode auch als schnellen Schalter und als Speicher mit Schaltzeiten von weniger als 1 ns. Die Tunneldiode wird auf einen ihrer stabilen Arbeitspunkte *A* oder *B* vorgespannt (8.1.3–12). Spannungsänderungen bewirken bei Erreichen von Punkt 1 ein nahezu trägheitsloses Springen auf Punkt 2 und im entsprechenden Fall von 3 auf 4. Trotz der physikalisch interessanten Wirkungsweise haben Tunneldioden bisher nur wenig Verbreitung gefunden.

Gunn-Dioden (Gunn-Elemente) sind im Sinne der hier beschriebenen Bauelemente keine Dioden, die zu ihrem Betrieb einen Übergang mit Gleichrichtereigenschaften benötigen. Für kleine elektrische Feldstärken, für die das Ohmsche Gesetz gilt, ist die mittlere Driftgeschwindigkeit der Ladungsträger der Feldstärke proportional. Die Proportionalitätskonstante ist die Beweglichkeit μ. Diese Proportionalität geht bei höheren Feldstärken verloren. Wird an eine homogene, mit sperrfreien Kontakten versehene Probe aus *n*-leitendem Galliumarsenid (GaAs) eine so hohe Gleichspannung angelegt, daß ihre Feldstärke ≈ 4 kV/cm überschreitet (J. B. GUNN, 1963), so setzen Stromschwingungen von einigen Gigahertz (GHz) ein. Die Stromschwingungen, deren Frequenz der Probenlänge indirekt proportional ist, werden durch schnell ($c \approx 10^7$ cm/s) über die Probenlänge wandernde Domänen hoher elektrischer Feldstärke hervorgerufen. Sie entstehen an der „Katode“ und verschwinden an der „Anode“. Ihr Zustandekommen wird mit Besonderheiten der Energiebandstruktur erklärt: Mit zunehmender Feldstärke erfolgt eine Umbesetzung von Elektronen (Elektronentransfer) von einem Leitungs-Teilband in ein energetisch höher gelegenes und durch eine geringere Elektronenbeweglichkeit ausgezeichnetes zweites Leitungs-Teilband. Diese negative Beweglichkeitsänderung $dV/dE < 0$ hat eine negative differentielle Leitfähigkeit zur Folge. Die außerordentlich kurzen Umbesetzungszeiten sind von der Größenordnung 10^{-12} s. Daraus ergeben sich Anwendungsmöglichkeiten in der Höchstfrequenztechnik und in der schnellen Digitaltechnik.

Metall-Halbleiter-Übergang (Schottky-Diode). Ein Metall-Halbleiter-Übergang stellt einen schichtförmigen, innigen Kontakt dar, der je nach den Eigenschaften seiner Bestandteile und den Herstellungsbedingungen entweder eine diodenartig nichtlineare oder eine Ohmsche Abhängigkeit zwischen Strom und Spannung zeigen kann. Die voneinander verschiedenen (Elektronen-) Austrittsarbeiten des Metalls W_M und des Halbleiters W_H verursachen eine Potentialdifferenz zwischen Metall und Halbleiter, die vom *pn*-Übergang (vgl. 8.1.3.1.) her bekannte Diffusionsspan-

nung U_D. Infolge des dadurch im Halbleiter bewirkten elektrischen Feldes werden an dessen Oberfläche Ladungen influenziert. Diese Verletzung der Neutralität führt zu einer Energiebandverbiegung im Halbleiter (Abb. 8.1.3–13). Die beiden möglichen Fälle für die Raumladungszone an der Grenzfläche sind Ladungsträgeranreicherung oder -verarmung. Ist die Austrittsarbeit des Metalls $W_M < W_H$, gelangen Elektronen leichter vom Metall in den Halbleiter als umgekehrt. Es entsteht an der Halbleitergrenzfläche eine Anreicherungsrandschicht, die eine lineare Strom-Spannungs-Kennlinie für beide Stromrichtungen aufweist. Diese Eigenschaften müssen in der Praxis bei allen Metall-Halbleiter-Kontakten gegeben sein, die als Stromzuführungen zu Bauelementen dienen sollen. Ist dagegen die Austrittsarbeit des Halbleiters $W_H < W_M$, gehen bevorzugt Elektronen vom Halbleiter in das Metall über. Die Halbleiterseite des Übergangs verarmt an Elektronen, sie lädt sich infolge der nun nicht mehr kompensierten Donatorionen positiv auf. An der Metallseite des Übergangs entsteht eine negative Oberflächenladung. Beim Anlegen einer äußeren Spannung ergibt sich eine Strom-Spannungs-Kennlinie mit Gleichrichtereigenschaften (Abb. 8.1.3–14). Bei der praktischen Herstellung von Schottky-Dioden ist die Abhängigkeit des Randschichttyps von der Differenz der Austrittsarbeiten nicht streng gültig, weil die schwierig vorhersagbaren Eigenschaften von Oberflächenzuständen das elektronische Verhalten des Metall-Halbleiter-Übergangs mitbestimmen. Schottky-Dioden haben gegenüber *pn*-Übergängen Vorteile im Verhalten beim Betrieb mit höchsten Frequenzen. Der Stromfluß wird nur von Majoritätsträgern (Elektronen im *n*-Gebiet) getragen, es gibt keine Injektion und Speicherung von Minoritätsträgern. Schottky-Dioden können daher als Schaltdioden mit Schaltzeiten unter 1 ns eingesetzt werden. In der Digitalelektronik, in der binäre Signale über logische Funktionen verknüpft werden, wird nach Vergrößerung der Arbeitsgeschwindigkeit (Taktfrequenz) gestrebt. In der Transistor-Transistor-Logik-Familie konnten unter Einbeziehung von Schottky-Dioden eine Verkürzung der Schaltzeit und eine Reduzierung der Verlustleistung erreicht werden.

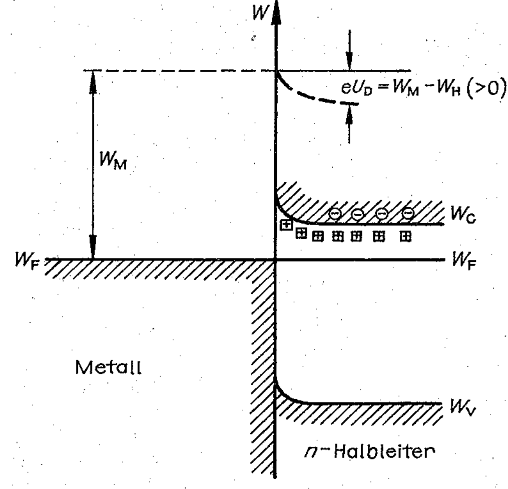


Abb. 8.1.3–13 Energiebandmodell des Übergangs Metall-*n*-Halbleiter mit Verarmungsrandschicht im stromlosen Zustand

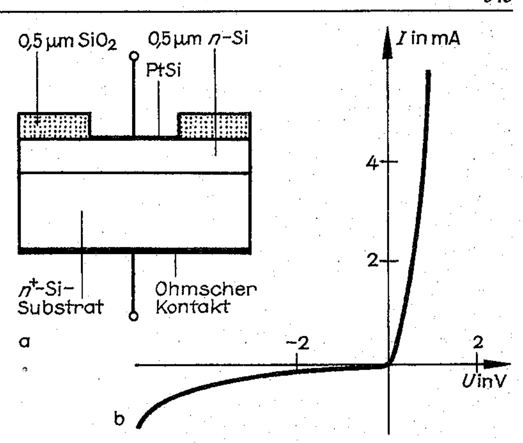


Abb. 8.1.3–14 Schottky-Diode, (a) Aufbau, (b) Strom-Spannungs-Kennlinie

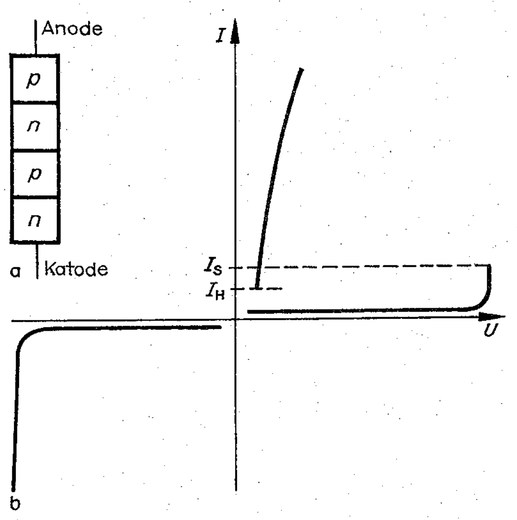


Abb. 8.1.3–15 Vierschicht-Diode, (a) Schichtfolge, (b) Strom-Spannungs-Kennlinie

Thyristoren können als Weiterentwicklung einer Vierschichtstruktur *pnpn*, der *Vierschichtdiode*, auch *Shockley-Diode*, (Abb. 8.1.3–15a) aufgefaßt werden. Liegt an der äußeren *p*-Schicht, die man als Anode bezeichnet, ein gegenüber der Katode K positives Potential, dann kann diese Schichtfolge zwei Betriebszustände annehmen, stellt also einen bistabilen Zweipol dar. Von den drei *pn*-Übergängen befinden sich der erste und dritte in Flußpolung und der mittlere in Sperrpolung. Der in diesem Zustand fließende Strom ist klein, es ist der von den in der jeweils inneren *n*- bzw. *p*-Schicht vorhandenen Minoritätsträgern herrührende Sperrstrom. Bei Erhöhung der Potentialdifferenz zwischen A und K kommt es zum Spannungsdurchbruch (vgl. Abb. 8.1.3–2), der infolge der hohen elektrischen Feldstärke am gesperrten *pn*-Übergang zur Erzeugung von Elektron-Defektelektron-Paaren führt. Damit setzt eine Ladungsträgerinjektion aus der äußeren *n*-Schicht ein. Diese Elektronen sind in der *p*-Seite des

Übergangs, an dem der Spannungsdurchbruch stattfindet, Minoritätsträger. Für sie hat der mittlere *pn*-Übergang Flußpolung. Durch die nun vorhandene hohe Ladungsträgerkonzentration kompensieren die beweglichen Ladungsträger die feste Raumladung und versetzen die Sperrschicht in den Flußzustand.

Die hohe Durchbruchfeldstärke bricht zusammen und bei Aufrechterhaltung des hohen Stromes geht die über die Diode abfallende Potentialdifferenz zurück. Für die Vierschichtdiode ergibt sich die in Abb. 8.1.3–15b dargestellte Strom-Spannungs-Kennlinie. Der niederohmige Zustand der Anordnung bleibt solange bestehen, solange der Strom die Haltestromstärke nicht unterschreitet. Da der Strom I_S bei eingeleitetem Spannungsdurchbruch größer ist, als der Haltestrom I_H, kommt ein Hystereseverhalten zustande. Die Polung der Spannung, bei der die beiden äußeren *pn*-Übergänge gesperrt sind, wird nicht bis zum Spannungsdurchbruch betrieben. Vierschichtdioden sind elektronische Schalter, die in der Anwendung gegenüber komplizierteren Bauelementen zu Schaltungsvereinfachungen führen.

Im Unterschied zur Vierschichtdiode hat der *Thyristor* einen herausgeführten Anschluß an die innere *p*-Schicht, den man als Steuerelektrode bezeichnet (Abb. 8.1.3–16a). Wie die Vierschichtdiode weist der Thyristor bei negativem Potential an der Anode A eine Sperrkennlinie im dritten

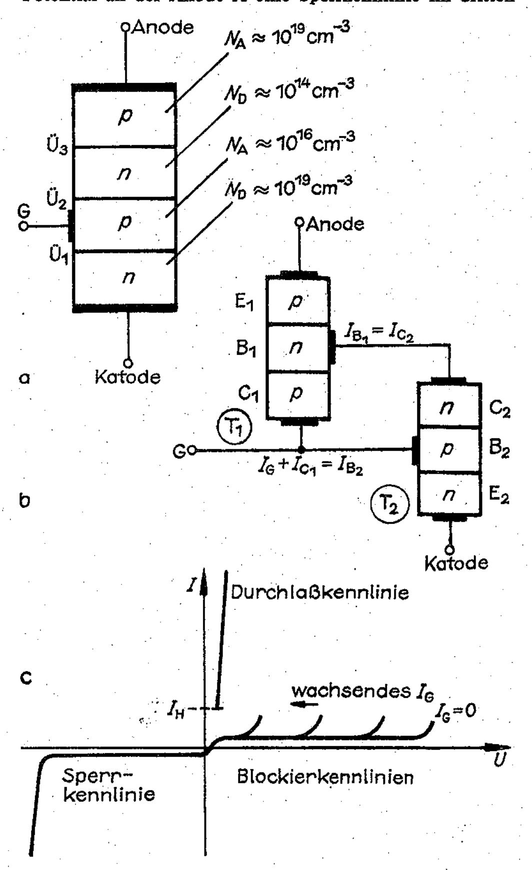


Abb. 8.1.3–16 Thyristor, (a) Schichtfolge, (b) Ersatzschaltung und (c) Strom-Spannungs-Kennlinien

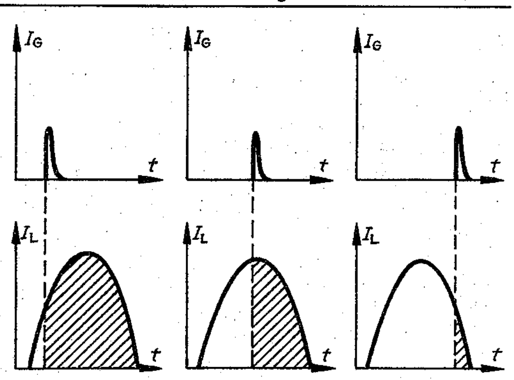


Abb. 8.1.3–17 Thyristor; Phasenschnittsteuerung mit Impulszündung zu unterschiedlichen, periodisch wiederkehrenden Zeiten

Quadranten auf, die von den beiden gesperrten Übergängen $Ü_1$ und $Ü_3$ herrührt. Liegt an der Anode positives Potential, so ergibt sich infolge des nunmehr gesperrten Übergangs $Ü_2$ die Blockierkennlinie im ersten Quadranten. Beim Erreichen des Spannungsdurchbruchs im Übergang $Ü_2$ bricht wie bei der Vierschichtdiode die Spannung an $Ü_2$ zusammen. Aus den hochdotierten äußeren Schichten werden die inneren Schichten mit beweglichen Ladungsträgern „zugeweht", wodurch die Raumladung, die für den Sperrzustand sorgte, kompensiert wird. Das Verhalten bei erneut ansteigender Spannung wird durch die Durchlaßkennlinie beschrieben. Bei dieser Polarität der äußeren Spannung hat der Thyristor wie die Vierschichtdiode zwei stabile Betriebszustände. Der Unterschied zur Vierschichtdiode besteht darin, daß der Übergang zum Durchlaßzustand im praktischen Betrieb nicht durch eine Erhöhung der Spannung zwischen Anode und Katode herbeigeführt wird, sondern durch einen Strom I_G, der der Steuerelektrode G zugeführt wird. Der Mechanismus des Übergangs vom Blockierzustand in den Durchlaßzustand wird verständlich, wenn man die Vierschichtstruktur des Thyristors in zwei zueinander komplementäre Transistoren (vgl. 8.1.3.3.) entsprechend Abb. 8.1.3–16b aufteilt. Danach ist der Kollektorstrom I_{C_1} des *pnp*-Transistors T_1 auch steuernder Basisstrom I_{B_2} des *npn*-Transistors T_2 und der Kollektorstrom I_{C_2} des Transistors T_2 steuert gleichzeitig den Transistor T_1 über dessen Basisstrom I_{B_1}. Wegen des zunächst gesperrten Übergangs $Ü_2$ ist der Basis-Emitter-Übergang von T_1 in Durchlaß- und der Basis-Kollektor-Übergang von T_1 in Sperrichtung gepolt. T_1 hat ebenso wie T_2 den Betriebszustand aktiv normal. Wird nun in die Steuerelektrode G, d. h. in die Basis des Transistors T_2, ein positiver Stromimpuls eingespeist, so fließen Defektelektronen in den Emitter von T_2. Damit wächst das Potential der Basis und löst die Injektion einer hohen Elektronenkonzentration in die Basis B_2 aus, die dort wegen der geringen Störstellenkonzentration mit nur unwesentlichen Rekombinationsverlusten als Minoritätsträger in den Kollektor von T_2 gelangen, und T_2 vergrößert seine Durchlässigkeit. Der Kollektorstrom I_{C_2}, der gleichzeitig Basisstrom I_{B_1} des Transistors T_1 ist, sorgt analog für eine verstärkte Injektion von Defektelektronen aus dem Emitter des Transistors T_1. Die injizierten Defektelektronen kommen über den Kollektor von T_1 zur Basis von T_2 und setzen dort den Prozeß fort, der mit dem äußeren Stromimpuls an der Steuerelek-

trode G, der Basis von T_2, eingeleitet wurde. Insgesamt entsteht eine lawinenartige Aufschaukelung des Stromes, nach deren Abschluß der mittlere *pn*-Übergang $Ü_2$ von beweglichen Ladungsträgern zugeweht worden ist. Damit ist die Spannung an $Ü_2$ zusammengebrochen und die Durchlaßkennlinie (Abb. 8.1.3–16c) erreicht. Im Verlauf des „Zündvorgangs" vergrößern sich die Stromverstärkungen in T_1 und T_2 von Werten unterhalb 0,1 bis auf mehr als 0,9. Das ist eine für das Funktionieren des Thyristors entscheidende Bedingung. Der Thyristor verbleibt solange im Durchlaßzustand, auch wenn über die Steuerelektrode kein Strom mehr fließt, bis die Stromstärke unter einen bestimmten Wert, den Haltestrom I_H, herabgesunken ist. Er braucht also im Gegensatz zum Transistor keine ständig vorhandene Steuergröße. Typische Zeiten für den Übergang in den gezündeten, d. h. Durchlaßzustand, sind $<10\,\mu s$, für den Übergang in den blockierenden Zustand $<100\,\mu s$, was eine Aussteuerung mit Frequenzen der Größenordnung 10 kHz ermöglicht. Die Leistung, die man im Thyristor zur Steuerung aufbringen muß, liegt um mehrere Größenordnungen unter der gesteuerten Leistung. Für Leistungsthyristoren betragen die Dauergrenzströme einige hundert Ampere, die Halteströme um 100 mA und die Zündspannungen wenige Volt. Damit ist der Thyristor als elektronischer Schalter das wichtigste Bauelement der Leistungselektronik. Die Anwendung von Thyristoren ist neben dem Einsatz als Gleich- oder Wechselrichter besonders für leistungsarme Regeleinrichtungen verbreitet. Häufig werden Thyristoren zur verlustarmen Helligkeitsregelung großer Beleuchtungsanlagen und zur Drehzahlregelung von Motoren genutzt.

Durch einen zum geeigneten Zeitpunkt periodisch in die Steuerelektrode eingespeisten Stromimpuls kann zu nahezu jedem gewünschten Zeitpunkt im Verlauf der positiven Halbwelle der Hauptstromkreis eingeschaltet werden (Phasenanschnittsteuerung Abb. 8.1.3–17). Mit zwei zueinander antiparallel geschalteten Thyristoren sind beide Halbwellen aussteuerbar. Einheitliche Bauelemente, die in beiden Stromrichtungen als elektronischer Schalter arbeiten, heißen Zweiwegthyristor (*Triac*). Thyristoren, bei denen das Einschalten durch Paarbildung infolge Lichteinstrahlung in den gesperrten Übergang geschieht, werden als *Fotothyristoren* bezeichnet.

8.1.3.3. Bipolar-Transistor. Aufbau und Wirkungsweise des Bipolartransistors. Bipolartransistoren bestehen aus zwei dicht benachbarten, sich gegenseitig beeinflussenden *pn*-Übergängen. Zu diesem Zweck wird die mittlere, beiden Übergängen gemeinsame Schicht vergleichsweise schmal gehalten (Abb. 8.1.3–18). Damit der Anteil der in die mittlere Schicht gelangenden Minoritätsträger, der dort auch rekombiniert, klein bleibt, muß ihre Diffusionslänge groß gegen die Schichtdicke sein. Das ist gleichbedeutend mit einer geringen Störstellenkonzentration der mittleren Schicht. Von den beiden Möglichkeiten der Schichtanordnung, *npn* oder *pnp*, wird die erste für Si-Transistoren, die zweite für Ge-Transistoren bevorzugt. Das Verhalten beider Transistortypen ist gleichartig. Da die einander entsprechenden Ladungsträger das Vorzeichen ändern, ist die Polarität der zwischen den Elektroden des Transistors liegenden Spannungen ebenfalls umzukehren. Die mittlere, schwach dotierte Schicht heißt Basis. Die äußeren Schichten sind stärker dotiert, am stärksten die als Emitter bezeichnete. Die verbleibende Schicht ist der Kollektor. Bipolartransistoren sind steuerbare, aktive Bauelemente, mit deren Hilfe eine Verstärkung elektrischer Signale möglich

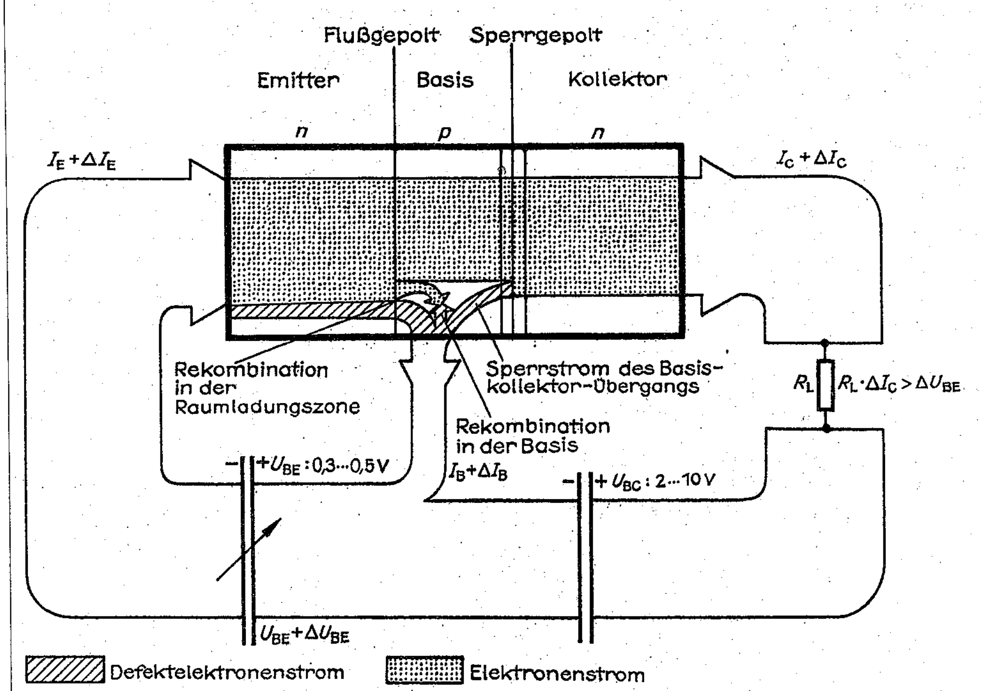


Abb. 8.1.3–18 Wirkungsweise des Bipolartransistors (schematisch) mit der Schichtfolge *npn* (Basisschaltung). Es werden nur Ladungsträgerströme betrachtet

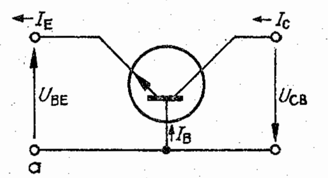
ist. Zur Erzielung eines Verstärkereffekts ist ein *pn*-Übergang in Fluß-, der andere in Sperrichtung zu polen. Ist der in Flußrichtung gepolte der Emitter-Basis-Übergang und der gesperrte der Basis-Kollektor-Übergang, so bezeichnet man den Betriebszustand des Bipolartransistors als „aktiv normal". Eine Vertauschung der Polungszuordnung (Betriebszustand „aktiv invers") bringt nichts prinzipiell Neues. Sind beide *pn*-Übergänge gesperrt bzw. in Durchlaßrichtung gepolt, ist keine Verstärkerwirkung möglich. Im aktiv normalen Betriebszustand fließt bei zunächst nicht beschaltetem Emitter über die gesperrte und daher hochohmige Kollektor-Basis-Diode der sehr kleine, temperaturabhängige und aus den Minoritätsträgern (Elektronen im *n*- und Defektelektronen im *p*-Gebiet) bestehende Sperrstrom. Eine geringe Flußspannung am Emitter-Basis-Übergang „emittiert" einen im Vergleich dazu großen, aus Elektronen bestehenden Injektionsstrom in das *p*-Gebiet der Basis. Durch die große Diffusionslänge für Elektronen im *p*-Gebiet rekombiniert nur ein geringer Anteil des Elektronenstroms. Ein weiterer geringer Teil des Stroms fließt im Emitter-Basis-Kreis. Der größte Teil des Stroms gerät als (dem Konzentrationsgradienten proportionaler) Diffusionsstrom durch die Basis an den Basis-Kollektor-Übergang. Da Elektronen in der Basis Minoritätsträger sind, können sich die Elektronen am Basis-Kollektor-Übergang „sammeln" und sie werden durch das höhere Potential an der Kollektor-Elektrode als Feldstrom abgesaugt. Betrachtet man eine kleine Änderung der Spannung zwischen Basis und Emitter ΔU_{EB}, so hat dies eine Emitterstromänderung ΔI_E zur Folge. Da $I_E \gtrapprox I_C$ gilt, ruft die Änderung ΔI_E eine fast gleichgroße Änderung ΔI_C hervor. Weil der Kollektor aber auf einem sehr viel höheren Potential liegt, verursacht eine kleine Änderung ΔU_{EB} am Lastwiderstand eine relativ große Kollektorspannungsänderung $\Delta U_{CB} \gg \Delta U_{EB}$, was einer Spannungsverstärkung entspricht. Im Gegensatz zu Elektronenröhren und Feldeffekttransistoren (vgl. 8.1.3.4.) muß bei Bipolartransistoren des niedrigen Eingangswiderstands wegen, hervorgerufen durch den in Flußpolung befindlichen Emitter-Basis-Übergang, eine Steuerleistung aufgebracht werden, da die im Eingangskreis fließenden Ströme nicht vernachlässigbar klein sind.

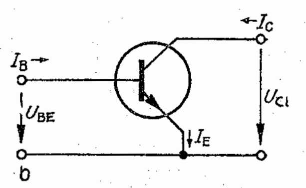
Grundschaltungen des Bipolartransistors. Der Transistor kann als Verstärkerbauelement in drei Grundschaltungstypen (Abb. 8.1.3–19) betrieben werden. Je nachdem, welche Transistorelektrode dem Eingang und dem Ausgang des Verstärkerbauelements gemeinsame Bezugselektrode ist, heißen die Grundschaltungen Basis-, Emitter- und Kollektorschaltung. Abb. 8.1.3–18 entspricht der Basisschaltung. Aus der Anordnung ist zu erkennen, daß der Eingangswiderstand wegen des in Durchlaßrichtung gepolten Emitter-Basis-*pn*-Übergangs klein ist. Der in Sperrichtung gepolte Basis-Kollektor-*pn*-Übergang bedingt einen großen Ausgangswiderstand. Da der Kollektorstrom um den Basisstrom kleiner ist als der Emitterstrom, ergibt sich eine Stromverstärkung von $\lessapprox 1$, dagegen ist die Spannungsverstärkung groß. Wegen des niedrigen Eingangswiderstands eignet sich diese Schaltung nur zur Verstärkung von Signalen, die aus Quellen mit kleinem Innenwiderstand R_i stammen.

Emitterschaltungen (Abb. 8.1.3–19b), bei der Eingangs- und Ausgangswiderstand näher beieinander liegen, werden am häufigsten verwendet, weil bei ihnen eine Zusammenkopplung von mehreren Stufen leichter ist, als bei der Basisschaltung. Diese Grundschaltung hat sowohl eine große Strom- wie auch Spannungsverstärkung, also auch eine beträchtliche Leistungsverstärkung. Ausgangs- und Eingangssignal sind gegeneinander um π in der Phase verschoben.

Kollektorschaltungen weisen einen hohen Eingangswiderstand und einen niedrigen Ausgangswiderstand auf. Die Stromverstärkung ist groß, die Spannungsverstärkung jedoch $\lessapprox 1$, womit sich diese Schaltung zur Anpassung von Signalquellen mit hohem Innenwiderstand an niederohmige Verbraucher eignet.

Kennlinienfelder des Bipolartransistors. Zur elektrischen Charakterisierung von Transistoren werden die gegenseitigen Abhängigkeiten der statischen Werte von Strömen und Spannungen in Kennlinienfeldern dargestellt (Abb. 8.1.3–20). Für den dynamischen Betrieb verhält sich der Transistor wie ein lineares Bauelement, wenn vorausgesetzt werden kann, daß die Änderungen von Spannungen und Strömen während des Betriebs klein gegenüber den statischen Werten sind (Kleinsignalbetrieb). Dann lassen sich die veränderlichen Eingangs- und Ausgangsströme und -spannungen durch lineare Gleichungen miteinander verknüpfen. Damit wird eine analytische Erfassung der Betriebseigenschaften von Transistorverstärkern möglich. Hilfsmittel hierbei ist die Vierpoltheorie, die allgemein der Untersuchung elektrischer Netzwerke dient. Daneben sind Verhaltensmerkmale des Transistors bei Aussteuerung in nichtlinearen Kennlinienbereichen, sein Impuls- und Schaltverhalten, sein Verhalten bei hohen Frequenzen, seine thermischen und Rauscheigenschaften für die praktische Anwendung interessant. Die Darstellung von Transistorkenngrößen in den Kennlinienfeldern geht davon aus, unter den drei Strömen I_E, I_B und I_C und den drei Spannungen U_{CE}, U_{CB} und U_{BE} eine solche Auswahl und wechselseitige Zuordnung zu treffen, daß für Anwendung ein Optimum an verwertbaren Informationen bereitsteht. Die jeweils dritte Größe (Strom bzw. Spannung) ist dabei durch ihren Zusammenhang mit den restlichen beiden bereits festgelegt. Für den Bipolartransistor hat sich, unabhängig davon, welche Grundschaltung (vgl. Abb. 8.1.3–19) vorliegt, die Darstellung von je vier Kennlinien bzw. Kennlinienfeldern durchgesetzt, die in vier Quadranten eines Achsenkreuzes eingeordnet werden (Abb. 8.1.3–20). Der erste Quadrant enthält das Ausgangskennlinienfeld mit der funktionalen Abhängigkeit des Ausgangsstroms I_2 von der Ausgangsspannung U_2. Parameter des Kennlinienfeldes ist der Eingangsstrom I_1. Der zweite Quadrant ist dem Übertragungskennlinienfeld vorbehalten. Ihm wird der Zusammenhang zwischen Eingangsstrom I_1 und Ausgangsstrom I_2 ent-





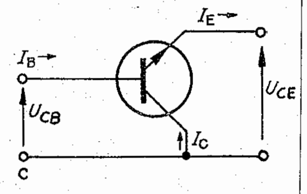


Abb. 8.1.3–19 Grundschaltungen des Bipolartransistors, (a) Basis-, (b) Emitter-, (c) Kollektorschaltung

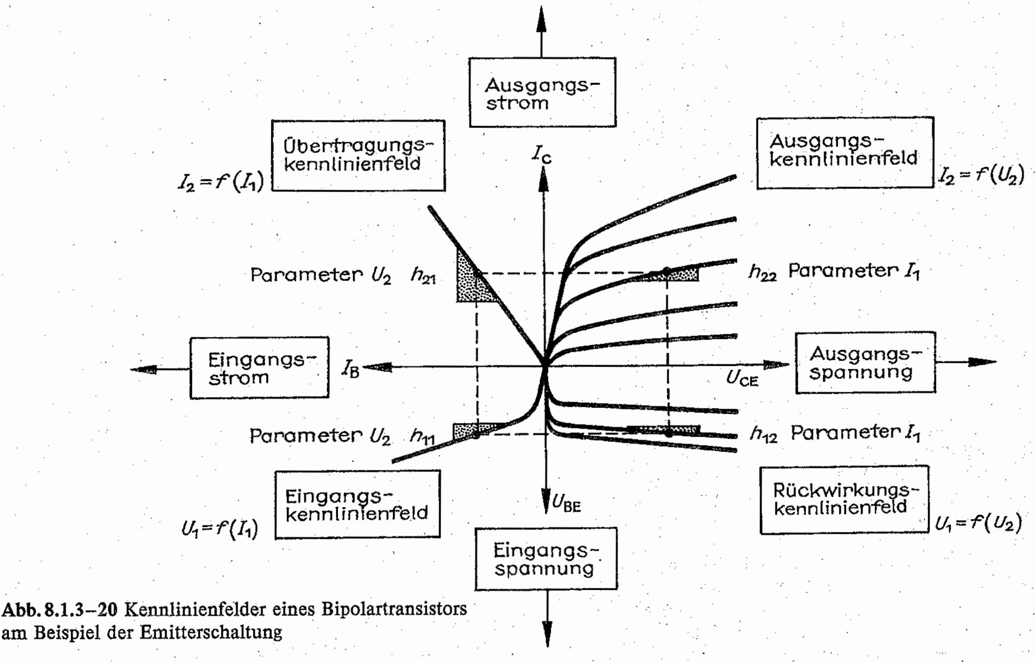


Abb. 8.1.3–20 Kennlinienfelder eines Bipolartransistors am Beispiel der Emitterschaltung

nommen. Parameter ist die Ausgangsspannung U_2. Im gewählten Beispiel der Emitterschaltung bedarf es nur einer Kennlinie, da die Abhängigkeit vom Parameter U_2 vernachlässigt werden kann. Der dritte Quadrant enthält das Eingangskennlinienfeld, die Abhängigkeit der Eingangsspannung U_1 vom Eingangsstrom I_1. Für die Emitterschaltung reicht wegen des geringen Einflusses von U_{CE} wieder die Darstellung einer Kennlinie. An der Exponentialform erkennt man das Verhalten der in Flußrichtung gepolten Emitter-Basis-Diode. Das Rückwirkungskennlinienfeld im vierten Quadranten stellt den Einfluß der Ausgangsspannung U_2 auf die Eingangsspannung U_1 dar. Wegen seiner geringeren Bedeutung wird es nicht immer angegeben.

Vierpolparameter des Bipolartransistors. Da sich der Transistor für kleine Spannungsänderungen u und kleine Stromänderungen i wie ein lineares Bauelement verhält, kann man jede der Grundschaltungen in Abb. 8.1.3–19 mit ihren zwei Eingangs- und zwei Ausgangskennlinien als einen Vierpol betrachten, dessen Eingangs- und Ausgangsströme und -spannungen durch zwei lineare Gleichungen verbunden sind. Welche Ströme oder Spannungen als unabhängige oder abhängige Variable betrachtet werden, ergibt sich aus der Forderung, daß die Konstanten der linearen Gleichungen, der Vierpolgleichungen, gut meßbar sein müssen, weil sie Charakteristika der entsprechenden Transistorschaltung darstellen. Als zweckmäßig für Bipolartransistoren im Niederfrequenzgebiet hat sich folgende Darstellung der beiden Vierpolgleichungen erwiesen:

$$u_1 = h_{11}i_1 + h_{12}u_2$$
$$i_2 = h_{21}i_1 + h_{22}u_2.$$

Da die Wechselströme und Spannungen an Eingang und Ausgang im Niederfrequenzgebiet den beiden Vierpolgleichungen genügen, enthalten die Vierpolparameter h_{ik}, die auch als Hybridparameter bezeichnet werden, typische Kenngrößen der Schaltung. Die Bedeutung der Parameter läßt sich aus einer Vierpolgleichung jeweils durch Nullsetzen der elektrischen Größe des anderen Vierpolparameters sofort ablesen. So ergibt sich:

$$h_{11} = \frac{u_1}{i_1}\bigg|_{u_2=0}$$ Eingangswiderstand bei Kurzschluß des Ausgangs,

$$h_{12} = \frac{u_1}{u_2}\bigg|_{i_1=0}$$ Spannungsrückwirkung bei Leerlauf des Eingangs,

$$h_{12} = \frac{i_2}{i_1}\bigg|_{u_2=0}$$ Stromverstärkung bei Kurzschluß des Ausgangs,

$$h_{22} = \frac{i_2}{u_2}\bigg|_{i_1=0}$$ Ausgangsleitwert bei Leerlauf des Eingangs.

Diese Aussagen können auf die drei Grundschaltungen des Bipolartransistors angewandt werden. Eingangs- und Ausgangsströme und -spannungen erfahren dann die der Schaltung entsprechende Zuordnung zu den Elektroden des Bipolartransistors. Die Anwendung auf die am häufigsten gebrauchte Emitterschaltung und die Verbindung zu den in Abb. 8.1.3–20 gegebenen Kennlinienfeldern ergibt sich dadurch, daß man den Wechselgrößen Änderungen der statischen Größen gleichsetzt. Da die h-Parameter, die reelle Größen sind, für verschiedene Grundschaltungen auch verschiedene Werte haben, wird für die Emitterschaltung ein Index hinzugesetzt:

$$h_{11_e} = \frac{\Delta U_1}{\Delta I_1}\bigg|_{\Delta U_2=0} = \frac{\Delta U_{BE}}{\Delta I_B}\bigg|_{\Delta U_{CE}=0}$$

$$h_{12_e} = \frac{\Delta U_1}{\Delta U_2}\bigg|_{\Delta I_1=0} = \frac{\Delta U_{BE}}{\Delta U_{CE}}\bigg|_{\Delta I_B=0}$$

$$h_{21_e} = \frac{\Delta I_2}{\Delta I_1}\bigg|_{\Delta U_2=0} = \frac{\Delta I_C}{\Delta I_B}\bigg|_{\Delta U_{CE}=0}$$

$$h_{22_e} = \frac{\Delta I_2}{\Delta U_2}\bigg|_{\Delta I_1=0} = \frac{\Delta I_C}{\Delta U_{CE}}\bigg|_{\Delta I_B=0}$$

Diese Größen können aus den Kennlinienfeldern bestimmt werden. Die Kennlinien im ersten bis vierten Quadranten liefern der Reihe nach die Vierpolparameter h_{22}, h_{21}, h_{12}, h_{11}. Für eine Anwendung von Bipolartransistoren bei höheren Frequenzen empfiehlt sich die Benutzung anderer Vierpolparameter, der Leitwertparameter y_{ik}. Für sie lauten die beiden Vierpolgleichungen

$$i_1 = y_{11}u_1 + y_{12}u_2$$
$$i_2 = y_{21}u_1 + y_{22}u_2$$

Die Vierpolparameter sind hier Leitwerte. Für höhere Frequenzen sind sie komplexe Größen. Die in der Vierpoltheorie angewandten mathematischen Methoden gestatten eine Umrechnung der Vierpolparametersätze von einer Darstellungsform in die andere. Auch innerhalb einer Darstellungsform ist für die h-Parameter z. B. aus dem gegebenen Parametersatz für die Emitterschaltung der Parametersatz für die Basisschaltung berechenbar. Daraus ist ersichtlich, daß ein Satz von Vierpolparametern das Bauelement umfassend kennzeichnet.

Betriebsgrößen des Bipolartransistors. Eine Erweiterung gegenüber der Betrachtung des offenen Transistorvierpols ist sein eingangsseitiger Abschluß mit dem aktiven Zweipol, der Signalquelle U_0 und deren Innenwiderstand R_q und der ausgangsseitige Abschluß mit dem passiven Zweipol, dem Verbraucher in Form des Lastwiderstands R_L (Abb. 8.1.3–21). Neben den zwei Vierpolgleichungen stehen nun noch zwei weitere Beziehungen zur Berechnung von Betriebsparametern zur Verfügung. Sie sind durch die in der Eingangs- und Ausgangsmasche vorhandenen Spannungen gegeben:

$$u_1 = -i_1R_a - u_0$$
$$u_2 = -i_2R_L$$

Aus den nun insgesamt vier Beziehungen sind die Betriebsgrößen des Transistorvierpols, d. h. die Eingangs- und Ausgangsströme und -spannungen, Eingangs- und Ausgangswiderstand, Strom-, Spannungs- und Leistungsverstärkung in Abhängigkeit von Last- und Quellwiderstand analytisch ausdrückbar. Tab. 8.1.3–22 enthält eine Auswahl.

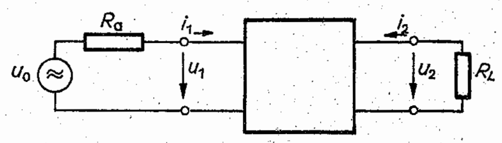


Abb. 8.1.3–21 Durch Signalquelle und Verbraucher abgeschlossener Transistorvierpol

Tab. 8.1.3–22 Betriebsgrößen des abgeschlossenen Transistorvierpols

Eingangswiderstand	$R_E = (h_{11} + \Delta h R_L)/(1 + h_{22}R_L)$
Ausgangswiderstand	$R_A = (h_{11} + R_q)/h_{22}R_q + \Delta h$
Stromverstärkung	$v_i = h_{21}/(1 + h_{22}R_L)$
Spannungsverstärkung	$v_u = -h_{21}R_L/h_{11} + \Delta h R_L)$
Leistungsverstärkung	$V_p = h_{21}^2 R_L/[(1 + h_{22}R_L)\,(h_{11} + \Delta h R_L)]$

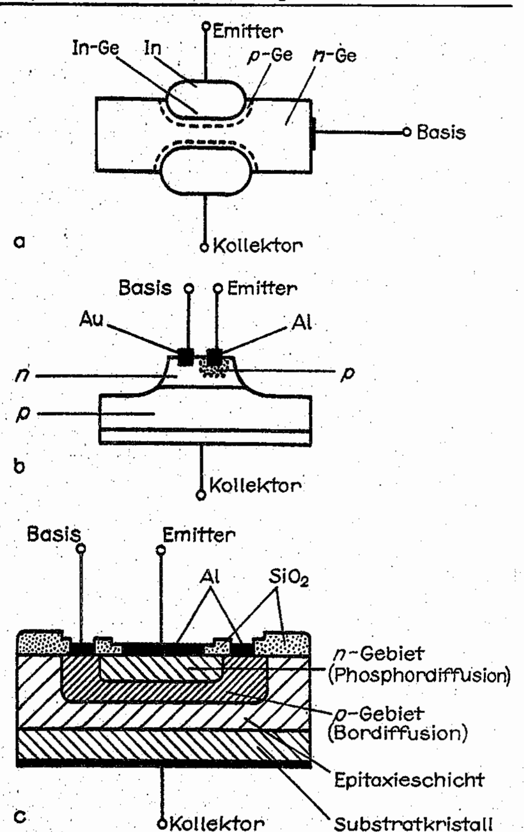


Abb. 8.1.3–23 Bauformen von Bipolartransistoren (schematisch), (a) Legierungstransistors, (b) Mesatransistor, (c) Epitaxie-Planar-Transistor

Zur Technologie der Herstellung von Bipolartransistoren. Die Technologie der Transistorherstellung hat seit der Entdeckung des Transistoreffektes viele Modifikationen durchgemacht. *Spitzentransistoren*, die aus den klassischen Kristalldetektoren hervorgegangen sind und an denen der Transistoreffekt 1948 erstmals nachgewiesen wurde, haben wegen ihrer Instabilitäten keine praktische Bedeutung mehr. Eine wichtige Etappe waren die *Legierungstransistoren*. Bei Temperaturen zwischen 600 und 700 °C werden nacheinander auf beiden Seiten eines n-leitenden Germaniumplättchens kleine Indiumkugeln einlegiert. An der Benetzungsfläche dringt Indium in das Germanium ein und erzeugt durch die Akzeptordotierung ein p-leitendes Gebiet und einen steilen Übergang der Störstellenkonzentration zum n-Gebiet. Legierungstransistoren (Abb. 8.1.3–23a) haben eine im MHz-Gebiet liegende relativ hohe Grenzfrequenz. Bei der tischartigen Bauform des *Mesatransistors*, einer weiteren technologischen Entwicklungsform des Bipolartransistors, wurde das Halbleiterplättchen nur einseitig bearbeitet. Durch Antimon-Diffusion wird eine n-leitende Schicht erzeugt. Mittels Lochmasken geschieht das Aufdampfen von Aluminium-Inseln, die nach dem Einlegieren das p-Gebiet des Emitters bilden. Es folgt die Kontaktierung der Basis und das Abätzen auf die Mesastruktur. Für sehr kleine Abstände zwischen Basis und Emitter werden

sehr hohe Grenzfrequenzen um 10 GHz erreicht (Abb. 8.1.3–23b). Bereits bei der Herstellung von Mesatransistoren wurde das aus der Kristallographie bekannte Epitaxieverfahren benutzt, um das wirksame Kollektorgebiet von dem nur als Zuleitung dienenden Bahngebiet des Germaniums zu trennen. Unter *Epitaxie* versteht man das Aufwachsen einer einkristallinen Schicht auf einer Kristallunterlage. Dabei wird die Kristallorientierung der Aufwachsschicht durch die Orientierung des Substrates bestimmt. Besonders gut wachsen unter geeigneten Bedingungen chemisch ähnliche Substanzen mit gleicher Kristallstruktur und möglichst nahe beieinanderliegenden Gitterkonstanten auf. Von Homoepitaxie wird gesprochen, wenn Substrat und Schicht, abgesehen von der Dotierung, aus dem gleichen Material bestehen, im Gegensatz zur Heteroepitaxie. Die gegenwärtig am meisten eingesetzten Transistoren sind *Silicium-Epitaxie-Planartransistoren* (Abb. 8.1.3–23c). Auf eine wenige Zehntel Millimeter dicke Scheibe aus gutleitendem, d. h. hochdotiertem *n*-Silicium, wächst in einem Epitaxieprozeß eine 10 bis 20 µm dicke hochohmige *n*-Siliciumschicht unter Beibehaltung der kristallografischen Orientierung auf. Über einen Oxidationsprozeß wird die Epitaxieschicht mit einer 1 µm-dicken Schicht von Siliciumdioxid SiO_2, einem sehr guten Isolator, überzogen. Das ist ein Vorzug des Halbleiterwerkstoffes Silicium, derart dünne, gut isolierende Schichten über einen technologisch relativ einfach zu beherrschenden Prozeß zu bilden. Die Isolationsschicht wird mit einem lichtempfindlichen Lack überzogen. Auf einer solchen Kristallscheibe von z. B. 10 cm Durchmesser werden gleichzeitig viele Transistoren hergestellt. Eine Maske, die die Konturen der Flächen enthält, an denen später die Oxidschicht abgeätzt werden soll, wird vor der Belichtung auf den Fotolack gelegt. Nach der Entwicklung wird zunächst das Fenster für die Basisdiffusion geätzt. Durch einen Diffusionsprozeß, meist mit Bor, entsteht im Basisgebiet *p*-Leitung. Nach der Basisdiffusion muß das Basisfenster durch Oxidation wieder verschlossen werden. Danach wird das Emitterfenster in der Isolierschicht angeätzt, es schließt sich eine Phosphordiffusion zur Herstellung des *n*-leitenden Emittergebiets an. Die unterschiedlich leitenden Gebiete des Transistors sind damit fertiggestellt. Für die Elektrodenkontaktierung werden die Kontaktfenster geöffnet und es wird Aluminium aufgedampft. Der Schutz der Oberfläche durch eine Oxidschicht ist ein während des gesamten technologischen Prozesses angewandtes Prinzip. Damit können durch Diffusion auf der Siliciumoberfläche auch wesentlich kompliziertere Strukturen erzeugt werden, was in der Halbleiterblocktechnik Voraussetzung für die Herstellung integrierter Schaltkreise ist.

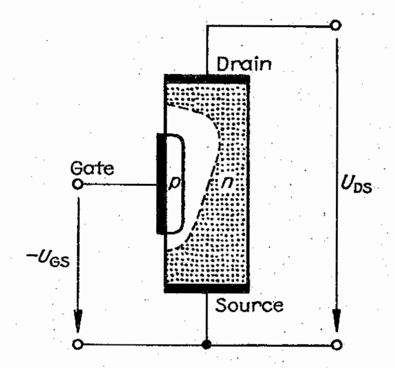


Abb. 8.1.3–24 Sperrschicht-Feldeffekttransistor (SFET), Einschnürung des Strompfades durch das Raumladungsgebiet des *pn*-Überganges

8.1.3.4. Feldeffekttransistor. Die Möglichkeit, auf einer Siliciumoberfläche ein beständiges und isolierendes Oxid aufzubringen, legte es nahe, ein Bauelement zu verwirklichen, bei dem ein Ladungsträgerstrom leistungslos durch die Isolationsschicht hindurch mit Hilfe der Influenzwirkung eines veränderlichen elektrischen Feldes gesteuert wird. Bauelemente dieser Art verbinden diesen Vorteil der Elektronenröhren mit dem des geringen Volumenbedarfs der Bipolartransistoren. Obwohl schon vor dem Bipolartransistor entworfen, erforderte die Herstellungstechnologie die stabile Beherrschung der Oberflächeneigenschaften des Halbleitermaterials, die z. B. beim Bipolartransistor lediglich als Störeinflüsse behandelt zu werden brauchten. Hierzu gehören u. a. Oberflächenzustände, hervorgerufen durch Adsorptions- und Desorptionserscheinungen. Daher konnten sich Feldeffekttransistoren erst später als Bipolartransistoren durchsetzen. Feldeffekttransistoren erfüllen funktionell die gleichen Aufgaben wie Bipolartransistoren und können in der Analogtechnik als Verstärker und in der Digitaltechnik als Schalter eingesetzt werden. Der für das Wirkprinzip erforderliche Strom besteht im Gegensatz zum Bipolartransistor nur aus einer Ladungsträgerart. Der Feldeffekttransistor ist daher ein Unipolartransistor. Für die Steuerung des Stroms bieten sich zwei Varianten.

Sperrschicht-Feldeffekttransistor. Im Sperrschicht-Feldeffekttransistor (Abb. 8.1.3–24) verläuft längs der Stromrichtung ein *pn*-Übergang. Diese ausschließlich in Sperrichtung gepolte Diode verändert in Abhängigkeit von der anliegenden Sperrspannung $-U_{GS}$ die Breite der von beweglichen Ladungsträgern entblößten Raumladungszone und steuert damit den Kanalquerschnitt für den Ladungsträgertransport. Die Elektroden des Feldeffekttransistors, die sperrschichtfrei kontaktiert sind, tragen die Bezeichnung Source (Quelle) und Drain (Senke), die Steuerelektrode heißt Gate (Tor). Steigt die Sperrspannung $-U_{GS}$ an, so vergrößert sich die Breite der Raumladungszone und der Kanalquerschnitt verengt sich. Da für Silicium die aus Minoritätsladungsträgern bestehenden Sperrströme sehr klein sind, wird mit der Spannungsänderung ΔU_{GS} praktisch leistungslos eine Stromänderung ΔI_D erzeugt. Die sich ergebenden Eingangswiderstände liegen bei $10^9\,\Omega$. In Abb. 8.1.3–24 soll die gestrichelte Linie im *n*-Gebiet die Ausdehnung der von beweglichen Ladungsträgern freien Raumladungszone des gesperrten *pn*-Übergangs andeuten. Die Unsymmetrie der Raumladungszone kommt bei fließendem Drainstrom I_D infolge des Spannungsabfalls von U_{DS}, der Spannung zwischen Drain und Source über der Kanallänge zustande. Das *p*-Gebiet an der Gate-Elektrode ist sehr hochdotiert, so daß sich die Sperrschicht vorwiegend in das geringer dotierte Gebiet des *n*-leitenden Kanals erstreckt. In Drain-Nähe ist die lokale Sperrspannung wesentlich größer, die Raumladungszone demzufolge auch weiter in das Kanalgebiet hinein ausgedehnt als in der Nähe der Source-Elektrode. Da der wirksame Widerstand des Kanalgebiets in der Nähe der Drain-Elektrode am größten ist, fällt dort auch der größte Teil der Spannung U_{DS} ab. Mit zunehmenden Werten von $-U_{GS}$ reduziert sich der Kanalquerschnitt und der Strom I_D wird nahezu unabhängig von U_{DS}. Diesen Abschnüreffekt des Stromkanals bezeichnet man als *pinch-off*. Die $I_D - U_{DS}$-Kennlinien (Parameter U_{GS}) weisen hier Sätti-

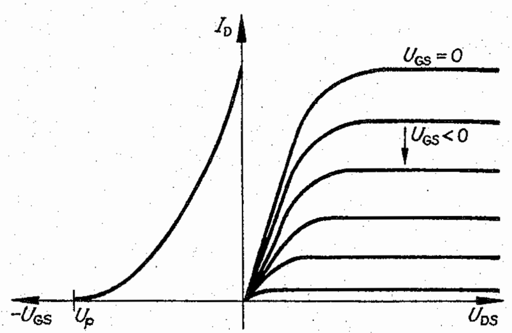


Abb. 8.1.3–25 Kennlinienfeld eines *n*-Kanal-Sperrschicht-Feldeffekttransistors

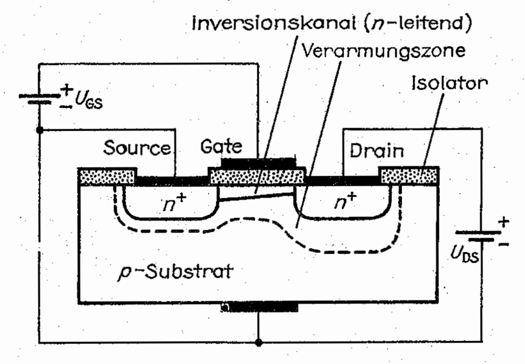


Abb. 8.1.3–26 MIS-Feldeffekttransistor vom *n*-Kanal-Verarmungstyp (Schnitt)

gungscharakter auf. Allerdings läßt sich die Drain-Source-Spannung nicht beliebig steigern, weil wegen ihres Einflusses auf die lokale Sperrspannung mit dem Spannungsdurchbruch des *pn*-Übergangs in Drain-Nähe gerechnet werden muß. Da der Eingangsstrom von Feldeffekttransistoren bedeutungslos ist, sind in zwei Kennlinientypen, der Transferkennlinie und den Ausgangskennlinien (Abb. 8.1.3–25), die wichtigsten Informationen über den Feldeffekttransistor enthalten. U_T ist die Schwellspannung, unterhalb derer der Kanal völlig abgeschnürt ist. Die Grenzfrequenz der Sperrschicht-Feldeffekttransistoren, die einige 100 MHz beträgt, hängt wesentlich von der Laufzeit der Ladungsträger im Kanal und damit von der Kanallänge ab, die wenige Mikrometer groß ist. Sperrschicht-Feldeffekttransistoren bieten den Vorteil, daß sie in integrierten analogen Schaltungen mit vertretbarem Mehraufwand mit Standard-Bipolartechnologien, und damit gemeinsam mit Bipolartransistoren, hergestellt werden können. Anstelle eines *pn*-Übergangs läßt sich die Steuerung auch mit einem Metall-Halbleiter-Kontakt (Schottky-Diode, vgl. 8.1.3.2.) realisieren (Metall-Halbleiter-Feldeffekttransistor [MESFET]).

MIS-Feldeffekttransistoren. Von größerer Bedeutung als die Sperrschicht-Feldeffekttransistoren sind die *Metall-Isolator-Silicium-Feldeffekttransistoren* (*MISFET*) (Abb. 8.1.3–26). Im Gegensatz zur Steuerung des Kanalquerschnitts bei den SFET wird bei den MISFET die Ladungsträgerkonzentration in einem Kanal an der Oberfläche mit Hilfe eines senkrecht zur Stromrichtung verlaufenden elektrischen Feldes gesteuert. Die Steuerelektrode ist dabei von der Kanaloberfläche isoliert. Neben den drei Anschlüssen für die Source-, Drain- und Gate-Elektrode wird zuweilen auch der Substratanschluß als Elektrode herausgeführt. Das elektrische Verhalten des Kanals an der Halbleiteroberfläche bestimmt sich aus Richtung und Stärke des elektrischen Feldes an der Gate-Elektrode. In Abb. 8.1.3–27 werden die Verhältnisse an einer MIS-Struktur mit *p*-leitendem Substrat im Bändermodell betrachtet. Bei *n*-leitendem Substrat kehren sich bei den entsprechenden Fällen die Polarität von U_{GS} und die Richtungen der Bandverbiegungen um. Die Energiedifferenz in der Lage der Fermi-Energien von Metall und Halbleiter ist der zwischen beiden anliegenden Spannung proportional. Mit W_i wird die Mitte zwischen der Energie der oberen Valenzband- und der Energie der unteren Leitungsbandkante gekennzeichnet. Im thermodynamischen Gleichgewicht entspricht W_i der Lage der Fermi-Energie bei eigenleitendem, also nichtdotiertem Halbleiter. *p*- und *n*-Leitung sind damit durch $W_F < W_i$ bzw. $W_F > W_i$ gekennzeichnet. Diese Aussage wird im Fall

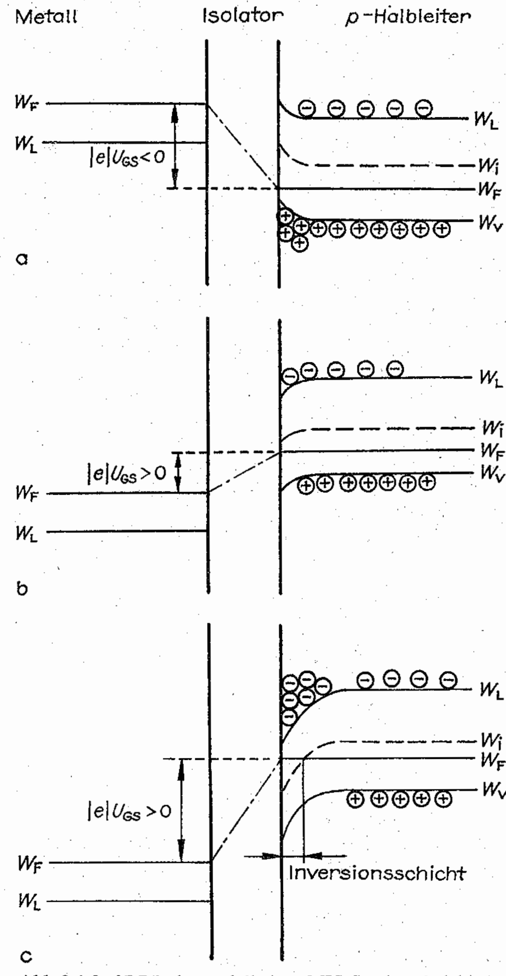
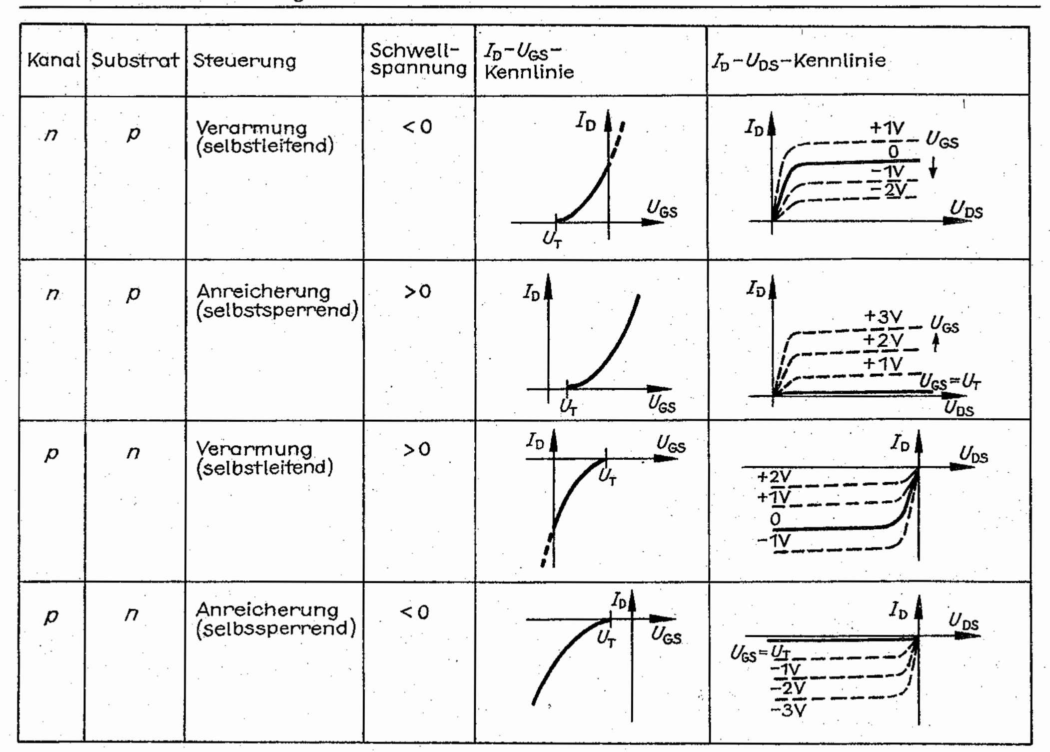


Abb. 8.1.3–27 Bändermodell einer MIS-Struktur bei (a) Anreicherung, (b) Verarmung, (c) Inversion



Kanal	Substrat	Steuerung	Schwellspannung	I_D-U_{GS}-Kennlinie	I_D-U_{DS}-Kennlinie
n	p	Verarmung (selbstleitend)	< 0		
n	p	Anreicherung (selbstsperrend)	> 0		
p	n	Verarmung (selbstleitend)	> 0		
p	n	Anreicherung (selbssperrend)	< 0		

Abb. 8.1.3–28 Typen von MIS-Feldeffekttransistoren

c benötigt. Im Verhalten des Kanals kann zwischen drei Fällen unterschieden werden. Der Anreicherungsfall a tritt ein, wenn bei p-leitendem Substrat das Gate gegenüber der Source-Elektrode negativ vorgespannt ist. Durch die Bandkantenverbiegung nähert sich die obere Kante des Valenzbandes W_v der Fermi-Energie W_F. Elektronen werden von der Oberfläche abgedrängt und Defektelektronen herangeführt. Die Majoritätsladungsträgerdichte, die Defektelektronendichte, übertrifft an der Halbleiteroberfläche diejenige im Innern des Halbleiters. Das führt zu einem Anwachsen der Leitfähigkeit in der Nähe der Oberfläche. Im Verarmungsfall b ist bei p-leitendem Substrat die Gate-Elektrode der Source-Elektrode gegenüber positiv vorgespannt. Die obere Valenzbandkante biegt sich von der Fermi-Energie weg und die Konzentration der Majoritätsladungsträger verringert sich. Damit weist die Oberfläche eine geringere Majoritätsladungsträgerdichte als das Halbleiterinnere auf, was einer verringerten Leitfähigkeit entspricht. Bei Verstärkung der positiven Vorspannung der Gate-Elektrode tritt der Inversionsfall c ein. Die Fermi-Energie gelangt infolge der stärkeren Bandverbiegung in die obere Hälfte der Bandlücke. Das entspricht einem lokalen Umschlag des Leitungstyps von der p-Leitung in die n-Leitung. Im Oberflächenbereich wird eine Inversionsschicht gebildet. Fehlt die Inversionsschicht, deren Ladungsträger thermisch gebildet werden müssen, gibt es keinen Stromfluß, da unabhängig von der Polarität von U_{DS} einer der beiden pn-Übergänge (vgl. Abb. 8.1.3–26) stets gesperrt ist. Zwischen der Inversionsschicht an der Halbleiteroberfläche und dem Halbleiterinnern liegt eine hochohmige, an beweglichen Ladungsträgern verarmte Zone, die die Inversionsschicht an der Oberfläche vom Halbleiterinnern abgrenzt. Aus den beim Sperrschicht-Feldeffekttransistor dargelegten Gründen ist die Verarmungszone der Nähe der Drain-Elektrode breiter als in der Nähe der Source-Elektrode. Die mit dem MISFET erzielten elektronischen Wirkungen beruhen auf der Steuerung der elektrischen Leitfähigkeit des Inversionskanals durch das von der Gate-Elektrode ausgehende elektrische Feld. Dabei sind Beiträge zu diesem Feld zu berücksichtigen, die nicht von einer äußeren Spannung herrühren. Hierzu zählen Oberflächenzustände, Ladungen an der Phasen-Grenze, Isolator-Halbleiter und die Differenz der Austrittsarbeiten. Für die Steuerung der Leitfähigkeit des Inversionskanals gibt es zwei Wege: 1. Wenn durch technologische Maßnahmen der Inversionskanal bereits ohne äußere Steuerspannung vorhanden ist, wird durch die Gate-Elektrode das Feld in einer solchen Weise geändert, daß sich die Ladungsträgerkonzentration im Kanal verringert, die Inversion also reduziert wird. Im gewählten Beispiel des MISFET mit p-Substrat entspricht das der Reduzierung der positiven Vorspannung zwischen Gate und Source. Diese Art der Steuerung heißt *Verarmungssteuerung*. Feldeffekttransistoren dieser Art, bei denen bereits bei $U_{GS} = 0$ ein wesentlicher Strom fließt, werden als „selbstleitend" bezeichnet. 2. Setzt die Entstehung des Inversionskanals und mit ihr das Anwachsen der Ladungsträgerkonzentration im Kanal erst mit anwachsendem äußeren elektrischen Feld durch Anlegen einer Spannung zwischen Gate- und Source-Elektrode ein, so spricht man von *Anreicherungssteuerung*. Da bei diesem Transistortyp bei $U_{GS} = 0$ kein Kanalstrom fließt, bezeichnet man ihn als „selbstsperrend". Beide Steuerungsarten können sowohl in n-leitenden wie

auch in p-leitenden Substraten verwirklicht werden. Es gibt daher vier Grundtypen von MIS-Feldeffekttransistoren, von denen aus technologischen Gründen n-Kanal-Verarmungs- und p-Kanal-Anreicherungstypen bevorzugt werden (Abb. 8.1.3–28). Der Leitungstyp des Kanals ist in jedem Fall invers zum Leitungstyp des Substrates. Bei herausgeführter Substratelektrode kann der MISFET auch vom Substrat aus gesteuert werden. Damit wird jedoch wie beim Sperrschicht-Feldeffekttransistor die Breite der Verarmungszone verändert.

Zur Herstellung von MISFET werden mit Hilfe der Fotolithographie Öffnungen zur Diffusion der Drain- und Source-Gebiete in die Isolationsschicht auf dem Siliciumsubstrat vorbereitet. Als Isolationsschichten kommen sowohl SiO_2 (MOSFET), Si_3N_4 (MNSFET), Al_2O_3 oder Kombinationen zur Anwendung. Die Gate-Elektroden bestehen aus Gold, Aluminium, oder polykristallinem Silicium (Silicium-Gate-Technologie). Die dünne Oxidschicht an der Gate-Elektrode ist empfindlich gegenüber elektrostatischen Aufladungen. Ein Durchschlagen wird durch einen Diodenschutz verhindert. Die MISFET-Technologie gestattet die Herstellung von Komplementärtransistoren (CMOS-Technik). Die Herstellung von MIS-Feldeffekttransistoren ist Ausgangspunkt einer Standardtechnologie von integrierten Schaltkreisen geworden, die vorwiegend in der Digitalelektronik eingesetzt werden. Gegenüber der Bipolartechnologie wird eine beträchtlich höhere Packungsdichte an Bauelementen erreicht, weil für eine Transistorstruktur in der MIS-Technologie wesentlich weniger Platzbedarf erforderlich ist, als in der Bipolartechnologie. Außerdem sinkt bei der Herstellung integrierter Schaltungen die Zahl der notwendigen Einzelprozeßschritte deutlich. Typische Kanallängen betragen $\approx 10\,\mu m$. Hohe Grenzfrequenzen fordern kurze Kanäle $< 1\,\mu m$. Je nach der Stärke des Drainstroms werden die Kanalbreiten zwischen 10 und 100 µm ausgelegt. Die Tiefe des Kanals ist einige 10 nm und die Dicke der Isolationsschicht zwischen Feldelektrode und Kanal ≈ 100 nm.

Für das Kleinsignalverhalten kann die Beschreibung der Eigenschaften von Feldeffekttransistoren durch Vierpolparameter in Leitwertform y_{ik} erfolgen:

$$i_1 = y_{11}u_1 + y_{12}u_2$$
$$i_2 = y_{21}u_1 + y_{22}u_2.$$

Wählt man die Source-Elektrode als die dem Ein- und Ausgang gemeinsame Bezugselektrode, so ergeben sich die für die Source-Schaltung typischen Eingangs- und Ausgangswechselgrößen:

$$i_G = y_{11}u_{GS} + y_{12}u_{DS}$$
$$i_D = y_{21}u_{GS} + y_{22}u_{DS}.$$

(i_G, i_D = Gate- bzw. Drainstrom, u_{GS}, u_{DS} = Gate-Source- bzw. Drain-Source-Spannung). Werden kleine Änderungen der statischen Größen als Wechselgrößen aufgefaßt, können Vierpolparameter den Kennlinien entnommen werden. Beispielsweise liefert der Anstieg der I_D-U_{GS}-Kennlinie den Leitwertparameter y_{21}, der der Steilheit S entspricht

$$\left.\frac{i_D}{u_{GS}}\right|_{u_{DS}=0} = \left.\frac{\Delta I_D}{\Delta U_{GS}}\right|_{U_{DS}=\text{const.}} = y_{21} = S$$

Die Steilheit von Feldeffekttransistoren ist merklich kleiner als beim Bipolartransistor. Dem Kehrwert des Ausgangsleitwerts

$$y_{22} = \left.\frac{i_D}{u_{DS}}\right|_{u_{GS}=0} = \left.\frac{\Delta I_D}{\Delta U_{DS}}\right|_{U_{GS}=\text{const.}}$$

entspricht der innere Widerstand R_i der Drain-Source-Strecke

$$R_i = \left.\frac{\Delta U_{DS}}{\Delta I_D}\right|_{U_{GS}=\text{const.}} = \frac{1}{y_{22}}$$

Wenn der Lastwiderstand $R_L \gg R_i$ ist, gilt für die Spannungsverstärkung v_u

$$v_u = R_i S = y_{21}/y_{22}.$$

Die Vierpolparameter müssen i. allg. als komplexe Größen aufgefaßt werden. Das wird besonders am Eingangsleitwert y_{11} deutlich. Das System Gateelektrode-Isolationssschicht-Kanal bildet einen Kondensator, dessen Einfluß bei höheren Frequenzen bestimmend wird.

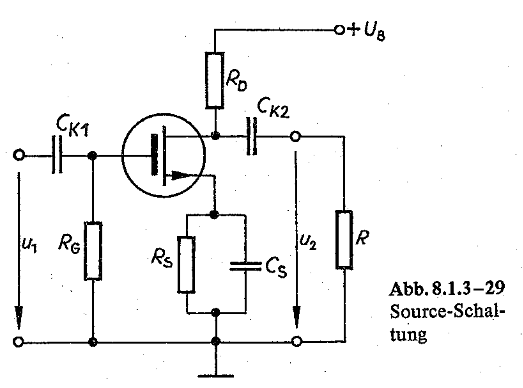


Abb. 8.1.3–29 Source-Schaltung

Die Source-Schaltung (Abb. 8.1.3–29) entspricht etwa der Emitterschaltung beim Bipolartransistor. Mit ihrer großen Spannungsverstärkung und einem hohen Eingangswiderstand, der bei niedrigen Frequenzen durch R_G gegeben ist, wird sie für Vorverstärker angewandt.

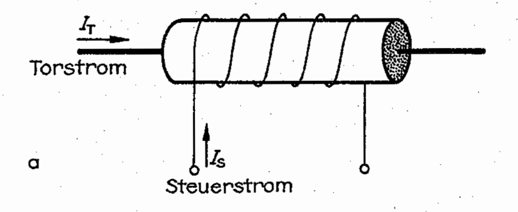
MIS-Feldeffekttransistoren werden in MIS-Schaltkreise integriert, in denen sie den einzigen Bauelementetyp darstellen.

8.1.4. Supraleitende Bauelemente

8.1.4.1. Zur Anwendung der Supraleitung. Im Gegensatz zu den Halbleiterbauelementen stellen bei supraleitenden Bauelementen Metalle im supra- und normalleitendem Zustand die stofflichen Träger der physikalischen Effekte dar. Die mit der Supraleitung verbundenen Eigenschaften bieten zahlreiche interessante Anwendungsmöglichkeiten. Der verschwindende elektrische Widerstand läßt Dauerstöße einer solchen Stärke zu, daß deren magnetische Wirkung eine schwebende und damit reibungsfreie Lagerung von Supraleitern, etwa im Kreiselkompaß ermöglicht. In supraleitenden Magneten erlaubt die verschwindende Joulesche Wärme Stromdichten anzuwenden, welche die in normalleitendem Kupfer um etwa drei Größenordnungen überschreiten. Auf diese Weise gelingt es, bei vergleichsweise wesentlich reduzierten Massen und geometrischen Dimensionen mit Supraleitern 3. Art stärkste Magnetfelder zu erzeugen. Es werden magnetische Induktionen von ≈ 20 T erreicht. Die durch die Supraleitung erzielbare extreme zeitliche Konstanz des Magnetfeldes und die realisierbare gute räumliche Homogenität haben auch neue physikalische Untersuchungsgebiete erschlossen. Die Erzeugung stärkster Magnetfelder ist andererseits für Fusionsreaktoren

des Tokamaktyps von Bedeutung, in denen thermonukleares Plasma extrem hoher Temperatur und Dichte durch eine magnetische Wand auf ein vorgegebenes Volumen zu begrenzen ist. Ebenso ermöglicht der Ersatz konventioneller durch supraleitende Magnete bei Teilchenbeschleunigern eine Steigerung der Teilchenenergien. Verbesserungen werden durch die Verwendung supraleitender Magnete auch bei Teilchendetektoren erzielt.
Zur Separation der beschleunigten positiven und negativen Ladungen in MHD-Generatoren werden ebenfalls supraleitende Magneten mit Erfolg benutzt.
Supraleitende Spulen können wichtige Bestandteile von mit Linearmotoren getriebenen Magnetkissenbahnen werden. Nach Projekten dieser Art wird bei hinreichend großer Geschwindigkeit durch supraleitende Spulen in der normalleitenden Schiene ein Strom induziert, dessen Kraftwirkung ausreicht, das Fahrzeug in der Schwebe zu halten. Für den Personentransport ist nach diesem Prinzip mit Geschwindigkeiten bis zu 500 km/h zu rechnen.
Für den Transport großer elektrischer Energiemengen können supraleitende Kabel bei geringem Volumenbedarf Leistungen der Größenordnung von Gigawatt übertragen. Den supraleitenden Kabeln der Starkstromtechnik sind supraleitende Band- oder Streifenleitungen der Informationselektronik gegenüberzustellen, durch die supraleitende Schaltungen verbunden werden können. An den Übergangsstellen Supraleiter – Normalleiter müssen dabei geeignete Bedingungen für die Energieein- und -auskopplung geschaffen werden. Die mit einer Einsparung von Elektroenergie und einem beträchtlich reduzierten Materialaufwand verbundene Anwendung der Supraleitung zum Transport von Elektroenergie und zur Erzeugung starker Magnetfelder haben die Suche nach Supraleitern aktiviert, deren Sprungtemperatur möglichst hoch liegt, am besten bei Raumtemperatur. Dieses Problem ist, obwohl intensiv bearbeitet, noch nicht gelöst.
Auch auf dem Gebiet der Informationselektronik werden Erscheinungen, die mit der Supraleitfähigkeit verbunden sind, angewandt. Dieses Teilgebiet der Elektronik trägt, der erforderlichen tiefen Temperaturen wegen, die Bezeichnung *Kryoelektronik*. Besonders anwendungsträchtig ist der gesteuerte Übergang vom supraleitenden zum normalleitenden Zustand. In einem als *Kryotron* (vgl. 8.1.4.2.) bezeichneten supraleitenden Bauelement wird dieser Übergang in einer supraleitenden Schicht durch die magnetische Wirkung eines Stromes, der in einem benachbarten zweiten supraleitenden Material fließt, hervorgerufen. Damit ist eine Schalterfunktion realisiert. Die Umschaltzeiten liegen je nach Bauart zwischen 10^{-3} und 10^{-8} s. Da die Umschaltung auch durch das Eigenmagnetfeld der umzuschaltenden Schicht selbst erfolgen kann, muß der Maximalstrom in der zu steuernden Strecke begrenzt werden. Mit Schaltelementen dieser Art können alle für die Datenverarbeitung nötigen logischen Funktionen aufgebaut und auch Speicherelemente realisiert werden.
Eine weitere Gruppe supraleitender Bauelemente bewerkstelligt unter Ausnutzung der auf dem Meißner-Ochsenfeld-Effekt beruhenden magnetischen Abschirmwirkung von Supraleitern Induktivitätsänderungen in supraleitenden Zweigen. Die Induktivitätsänderungen können mehr als zwei Größenordnungen betragen. Der induktivitätsarme Zustand eines (ständig supraleitenden) Torleiters, der sich zwischen zwei durch Isolierschichten von ihm getrennten Steuerleitern befindet, wird im supraleitenden Zustand der Steuerleiter erzielt.
Der Zustand großer Induktivität des Torleiters ist mit dem normalleitenden Zustand der Steuerleiter verknüpft. Schließlich sind kryoelektronische Bauelemente anzuführen, in denen an supraleitenden Materialien ein Tunneleffekt (vgl. 8.1.3.2.) stattfindet. Hier ist zu unterscheiden zwischen dem „normalen" Einelektronen-Tunneleffekt, wie er an der Tunneldiode (vgl. 8.1.3.2.) vorliegt und beispielsweise auch zwischen einem Normalleiter und einem davon durch eine dünne isolierende Schicht getrennten Supraleiter auftritt einerseits und dem Tunneleffekt von Cooper-Paaren zwischen zwei durch eine dünne isolierende Schicht voneinander getrennten Supraleitern (Josephson-Effekt) andererseits. Anordnungen mit vor sich gehendem Tunneleffekt von Cooper-Paaren bezeichnet man als Josephson-Tunnelelemente. Zwei über supraleitende Induktivitäten miteinander verbundene Josephson-Tunnelelemente bilden ein supraleitendes Quanten-Interferenz-Bauelement (SQUID). In diesem Bauelement treten bei geringsten Änderungen des magnetischen Flusses Interferenzen des Josephson-Tunnelstromes auf. Damit konnten die bisher empfindlichsten Messungen des magnetischen Flusses Φ bis auf Bruchteile des Elementarquantums des magnetischen Flusses $\Phi_0 = h/(2e) = 2 \cdot 10^{-15}$ Wb durchgeführt werden. Änderungen der magnetischen Induktion B von 10^{-12} T sind noch nachweisbar. Die hieraus ableitbaren strom- bzw. spannungsproportionalen Größen lassen die Verwendung dieser supraleitenden Anordnungen als empfindliche Galvanometer zu.

8.1.4.2. Kryotron. Das Kryotron nutzt den Sachverhalt, daß der Übergang vom supraleitenden zum normalleitenden Zustand auch bei gleichbleibender Temperatur, aber



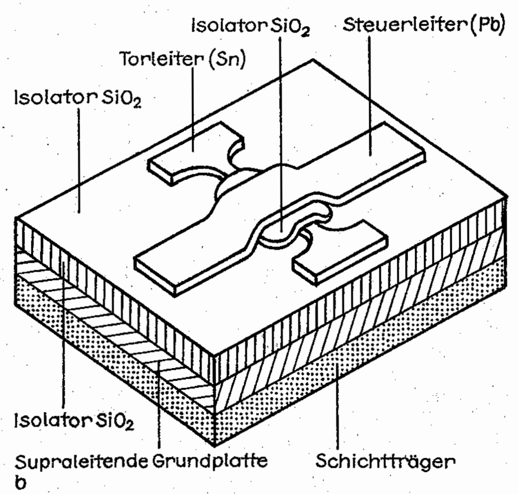


Abb. 8.1.4–1 (a) Drahtkryotron, (b) Kreuzkryotron

bei Vorhandensein eines Magnetfeldes vollzogen werden kann. Die zum Übergang nötige kritische Feldstärke H_c ist angenähert $H_c = H_0[1 - (T/T_c)^2]$. Man kann das Kryotron als Vierpol betrachten. In seinem Eingangskreis, in dem der Steuerleiter liegt, ruft ein fließender Strom ein Magnetfeld hervor, unter dessen Einfluß der im Ausgangskreis befindliche Torleiter vom supraleitenden in den normalleitenden Zustand übergeht. Die erste Ausführung eines Kryotrons, das Drahtkryotron (Abb. 8.1.4–1a), besteht aus einem Tantaldraht von $\approx 0{,}3$ mm Durchmesser und einer Länge von 10 bis 20 mm, um den ≈ 100 Windungen Niobiumdraht von 0,05 mm Durchmesser gewickelt sind. Die kritische Feldstärke für Tantal ist bei der Temperatur des flüssigen Heliums klein gegen die des Niobiums, so daß der Übergang des Tantaldrahts von supraleitenden zum normalleitenden Zustand mit hinreichend kleinen Stromstärken vollzogen werden kann. Die Zeitdauer des Übergangs wird von der Zeitkonstanten $\tau = L/R_n$ der Anordnung bestimmt. L ist die Induktivität der Niobiumdrahtwicklung und R_n der Widerstand des Tantaldrahts im normalleitenden Zustand. Für das Drahtkryotron ergeben sich relativ große Zeitkonstanten zwischen 10^{-4} und 10^{-3} s. Wesentlich kürzere Schaltzeiten von 10^{-7} bis 10^{-8} s erhält man mit Kryotrons aus supraleitenden dünnen Schichten. Ein solches Dünnschichtelement ist das Kreuzkryotron (Abb. 8.1.4–1b).
Der Steuerleiter befindet sich ständig im supraleitenden Zustand, während der Torleiter als aktives Element sowohl den supraleitenden als auch den normalleitenden Zustand annimmt. Um die kritische Stärke des Magnetfelds H_c klein zu halten, wird die Arbeitstemperatur nahe der kritischen Temperatur gewählt. Die Schichtdicken des Torleiters betragen einige hundert Nanometer. Im Zwischenzustand zwischen supra- und normalleitendem Zustand hat der Torleiter einen großen differentiellen Widerstand. Wählt man den Arbeitspunkt in diesem Zwischenzustand, so läßt sich das Kreuzkryotron auch als Kleinsignalverstärker verwenden.

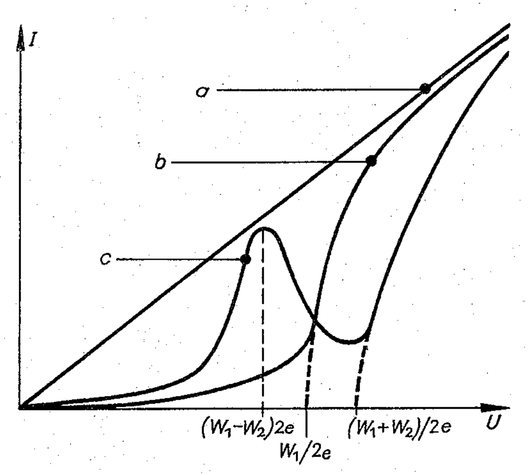


Abb. 8.1.4–2 Tunneleffekt ungepaarter Elektronen zwischen zwei durch eine Isolationsschicht getrennten Metallen M_1, M_2, (a) beide Metalle normalleitend ($T_{c_1}, T_{c_2} < T$), (b) M_1 supraleitend ($T_{c_1} < T < T_{c_2}$), (c) beide Metalle supraleitend ($T < T_{c_1}, T_{c_2}$). W_1, W_2 = Größe der Energielücken der Metalle M_1, M_2 im supraleitenden Zustand. Für $T = 0$ nehmen (b) und (c) den punktierten Verlauf, Stromfluß setzt dann erst bei den Minimalspannungen $W_1/2e$ bzw. $W_1 + W_2/2e$ ein

8.1.4.3. Tunnelelemente mit Tunnelung ungepaarter Elektronen. Der Einelektron-Tunneleffekt zwischen normal- und/oder supraleitenden Metallen hat Ähnlichkeiten mit dem Tunneleffekt in Halbleitern. Es treten analog zur Halbleitertunneldiode (vgl. 8.1.3.2.) deutbare Erscheinungen auf. Die zwischen beiden Metallen für die Tunnelung nötige Isolatorschicht hat eine Dicke von 2 bis 10 nm. Als Metall-Isolator-Metall-Systeme werden Aluminium/Aluminiumoxid/Blei, Aluminium/Aluminiumoxid/Zinn oder auch Zinn/Zinnoxid/Blei benutzt. Die Sprungtemperaturen T_c betragen für Blei 7,23 K, für Zinn 3,721 K und für Aluminium 1,175 K. Summarisch sind drei Zustandskombinationen einstellbar: 1. Beide Metalle sind normalleitend, 2. ein Metall ist normalleitend, das andere supraleitend, 3. beide Metalle sind supraleitend. Alle drei Fälle lassen sich mit einem Tunnelelement bei verschiedenen Arbeitstemperaturen einstellen. Als typische Strom-Spannungs-Kennlinien ergeben sich die in Abb. 8.1.4–2 dargestellten Verläufe. Ihr Zustandekommen erklärt sich aus dem Schema der Energiebandstruktur an der Tunnelschicht Abb. 8.1.4–3.
Abb. 8.1.4–3 enthält in den dargestellten Diagrammen die Energiebänder der beiden, an der isolierenden Zwischenschicht angrenzenden Metalle. Die vertikale Richtung ist ein Maß für die Energie W der Einelektronenzustände, in die horizontale Richtung ist die Zustandsdichte $Z(W)$ im jeweiligen differentiellen Energieintervall aufgetragen. Im Fall zweier an die isolierende Schicht grenzender normalleitender Metalle (Abb. 8.1.4–3a) liegt im thermodynamischen Gleichgewicht in beiden normalleitenden Metallen die Fermi-Energie W_F auf gleicher Höhe.
Da W_F bei $T = 0$ die Energie ist, für die Zustände mit $W < W_F$ alle besetzt und Zustände mit $W > W_F$ alle unbesetzt sind, kann kein Strom fließen. In Abb. 8.1.4–3 sind besetzte Zustände durch horizontale Schraffur dargestellt. Durch Anlegen einer äußeren Spannung U werden die Fermi-Energien beider Metalle gegeneinander verschoben. Es fließt ein Strom, da nun besetzten Zuständen in dem einen Metall unbesetzte Zustände gleicher Energie im anderen Metall gegenüberstehen. Im Gegensatz zur Halbleitertunneldiode (vgl. 8.1.3.2.) ergibt sich beim Tunneleffekt normalleitender Metalle infolge der anderen Bandstruktur eine lineare Strom-Spannungs-Kennlinie (Abb. 8.1.4–2a).
Grenzen ein Supraleiter und ein normalleitendes Metall an die Isolationsschicht, ergeben sich die in Abb. 8.1.4–3b dargestellten Verhältnisse. Der Stromfluß vom Normalleiter zum Supraleiter kann erst dann einsetzen, wenn die angelegte Spannung auf den Wert $W_1/2e$ angewachsen ist. Es ist ein steiler Anstieg des Stromes zu erwarten, da ausreichend unbesetzte Zustände im Supraleiter vorhanden sind. Für $T = 0$ setzt der Strom erst bei der Spannung $W_1/2e$ ein. Die in Abb. 8.1.4–2b dargestellte Kennlinie ist für $T > 0$ gültig. Die scharfe Grenze zwischen besetzten und unbesetzten Zuständen verwischt sich mit steigender Temperatur und bei $T > 0$ sind auch oberhalb W_F im Normalleiter schon besetzte Zustände vorhanden, die dann wie in Abb. 8.1.4–2b dargestellt, bereits bei kleineren Spannungen als $W_1/2e$ einen Beitrag zum Tunnelstrom liefern.
Zwischen zwei Supraleitern mit unterschiedlich großen Energielücken $W_1 + W_2$ kann bei $T = 0$ erst dann ein Strom fließen, wenn besetzte und unbesetzte Zustände einander in beiden Supraleitern gegenüberstehen. Das ist dann der

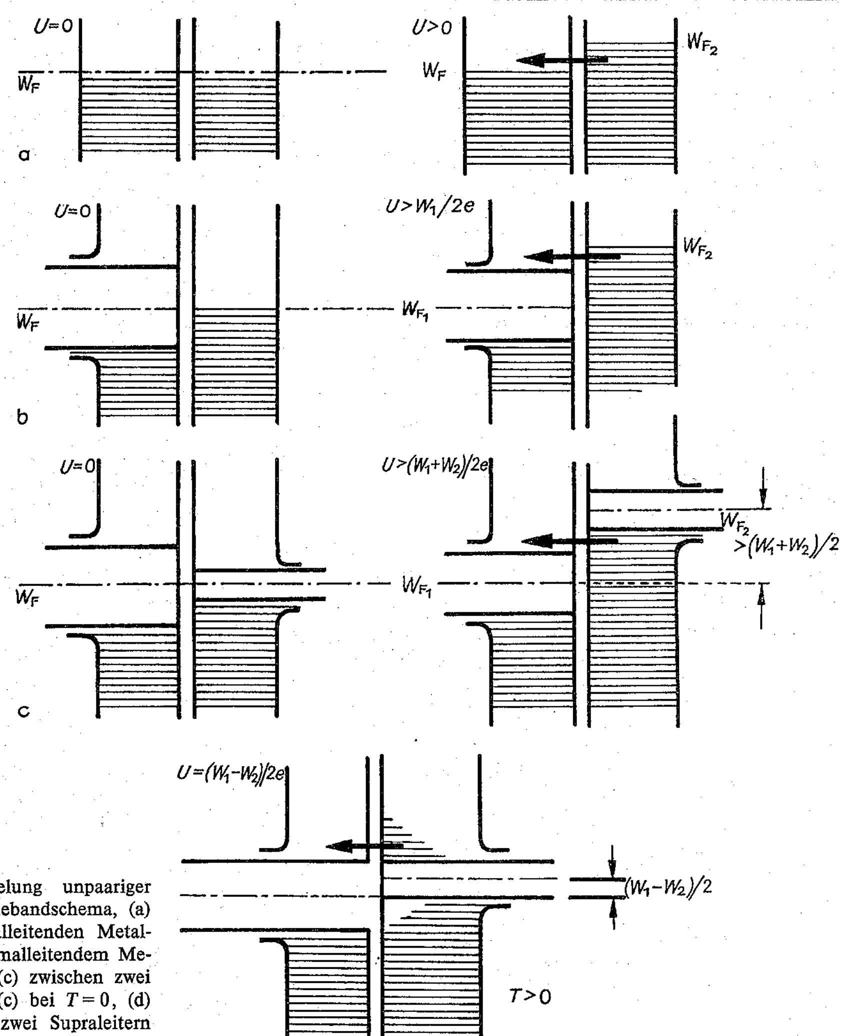


Abb. 8.1.4–3 Tunnelung unpaariger Elektronen im Energiebandschema, (a) zwischen zwei normalleitenden Metallen, (b) zwischen normalleitendem Metall und Supraleiter, (c) zwischen zwei Supraleitern; (a) bis (c) bei $T=0$, (d) Tunnelung zwischen zwei Supraleitern bei $T>0$

Fall, wenn an das System eine Spannung $U \geqq [(W_1 + W_2)/2]/e$ gelegt wird. Die Strom-Spannungs-Kennlinie in Abb. 8.1.4–2 würde, von kleinen Spannungen kommend, mit dem punktierten Verlauf beginnen.

Der experimentell beobachtete Verlauf der ausgezogenen Kennlinie in Abb. 8.1.4–2c läßt sich im Energiebandschema (Abb. 8.1.4–3d) verstehen. Für $T>0$ sind in dem Supraleiter mit der kleineren Energielücke bereits Elektronen durch thermische Anregung über die verbotene Zone der Energielücke gelangt und haben dort Zustände besetzt. Unterhalb der Energielücke sind Defektelektronen (Löcher) zurückgeblieben.

Wird das Tunnelelement mit einer Spannung von $U = [(W_1 - W_2)/2]/e$ vorgespannt, stimmen die Oberkanten der Energielücken überein und für die Tunnelung der thermisch angeregten Elektronen ergeben sich die günstigsten Bedingungen. Das ist die Erklärung für das relative Maximum bei der Spannung $U = [(W_1 - W_2)/2]/e$ in der Kennlinie Abb. 8.1.4–2c. Die Tunnelelemente, die auch praktisch angewandt werden, sind damit gleichzeitig zum Nachweis der Energielücke in Supraleitern herangezogen worden. Die Breite der Energielücke beträgt bei den Supraleitern wenige Millielektronenvolt im Gegensatz zur Halbleitertunneldiode, bei der die Energielücke in der Größenordnung von Elektronenvolt liegt. Die Breite der Energielücke von Supraleitern ist temperaturabhängig. Sie verschwindet bei der Sprungtemperatur T_c des Supraleiters. Die Zustände, die sich beim normalleitenden Metall bei Energien befinden, die im Supraleiter zur Energielücke zählen, lagern sich in unmittelbarer Umgebung der Energielücke an. Das ist in Abb. 8.1.4–3 schematisch durch die sich in der Nähe der Lücke vergrößernde Zustandsdichte dargestellt. Die Darstellung von Transportphänomenen im Energiebänderschema ist nur für Teilchen möglich, die der Fermi-Statistik unterliegen. Die für die Supraleitung maßgeblichen Elektronenpaare unterliegen jedoch der Bose-Statistik. Sie

können in diese Darstellung nicht einbezogen werden. Die Energiedifferenz der Bandlücke W_L entspricht der Bindungsenergie der Elektronenpaare. Elektronen mit diesen Energien bilden die Cooper-Paare, sie tauchen also in der Einelektronendarstellung des Bänderschemas nicht auf.
Der fallende Kennlinienanteil in Abb. 8.1.4–2c zwischen $[(W_1 - W_2)/2]/e < U < [(W_1 + W_2)/2]/e$ bedeutet einen negativen differentiellen Widerstand. Für ein Tunnelelement der Zusammensetzung $Al/Al_2O_3/Pb$ tritt der negative differentielle Widerstand bei ≈ 1 mV in einem Intervall von wenigen Zehntel Millivolt bei Strömen von 50 µA auf. Mit Tunnelelementen sind Verstärker, Schalter und Oszillatoren entwickelt worden.

8.1.4.4. Josephson-Tunnelelemente. Die Josephson-Effekte sind Erscheinungen, die bei der Tunnelung von Cooper-Paaren zwischen zwei durch eine ≈ 2 nm, d. h. etwa fünf Atomlagen dicke, isolierende Schicht (Abb. 8.1.4–4) ge-

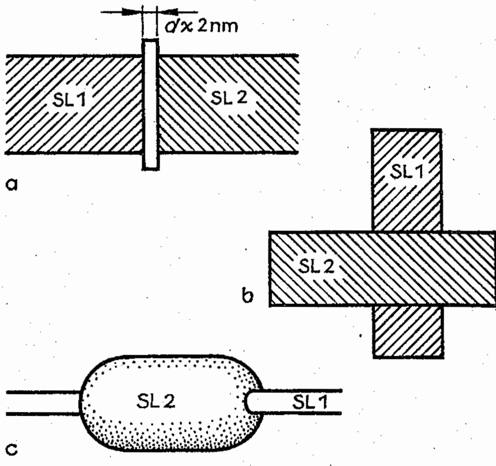


Abb. 8.1.4–4 Arten von Josephson-Übergängen, (a) schematisch, (b) supraleitende dünne Schichten, der Übergang ist oxidiert, z. B. Sn/SnO/Sn, (c) supraleitender oxidierter Draht, z. B. Nb, und supraleitender Löttropfen, z. B. Sn

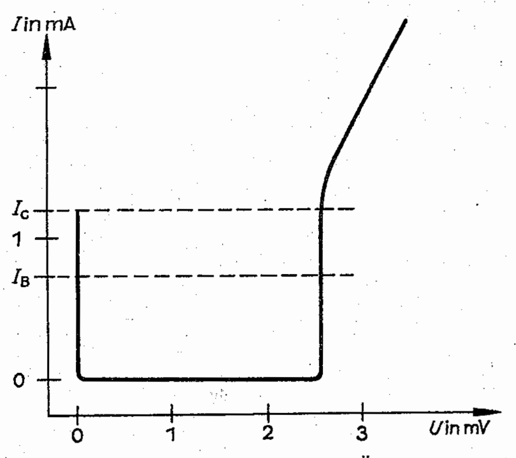


Abb. 8.1.4–5 Kennlinie eines Josephson-Überganges, I_c = kritische Stromstärke

trennte Supraleiter auftreten. Die isolierenden Schichten zwischen den beiden Supraleitern werden als Josephson-Übergänge (vgl. Abb. 8.1.4–4) bezeichnet. Systeme dieses Aufbaus haben die Eigenschaft, daß durch sie verlustlos ein Strom fließen kann, der nach den durch die Parameter des Systems gegebenen Werten zwischen $\pm I_c$ liegen kann. Dieser Zustand bleibt solange erhalten, bis die von außen angelegte Spannung U ausreicht, daß die Energielücken der Supraleiter von ungepaarten Elektronen überwunden werden kann. Das ist oberhalb des kritischen Stromes I_c der Fall. Am Josephson-Übergang tritt dann ein endlicher Spannungsabfall auf, der durch den im Stromkreis liegenden äußeren Widerstand beeinflußt wird. Bauelemente, denen Josephson-Übergänge zugrunde liegen, werden als Josephson-Tunnelelemente bezeichnet. Die Strom-Spannungs-Kennlinie eines Josephson-Tunnelelements zeigt Abb. 8.1.4–5. Das beschriebene Phänomen wird als *Gleichstrom-Josephson-Effekt* bezeichnet. Es wird ersichtlich, daß mit der gegebenen Umschaltungsmöglichkeit vom supraleitenden zum normalleitenden Zustand eine Speicherzelle aufgebaut werden kann. Die Umschaltung vom supraleitenden Zustand, in dem ein Strom I_B eingestellt sei, in den normalleitenden wird durch einen Zusatzstrom I_z bewirkt, der so gewählt ist, daß $I_B + I_z > I_c$ wird. Das Tunneln der Cooper-Paare geht bei Strömen $> I_c$ in das Tunneln ungepaarter Elektronen über. Dadurch entsteht am Josephson-Übergang ein Spannungssprung. Eine andere Umschaltmöglichkeit, um vom supraleitenden zum normalleitenden Zustand zu gelangen, ist die Verwendung eines Steuerleiters analog zum Kryotron. Der Tunnelübergang stellt den Torleiter dar, der vom Magnetfeld des Steuerleiters beeinflußt wird. Im Gegensatz zum Kryotron bleiben beim Schaltvorgang alle Teile des Josephson-Übergangs supraleitend. Anordnungen dieser Art heißen auch, je nachdem ob der Steuerleiter parallel oder senkrecht zum Torleiter verläuft, Tunnelparallel- oder Tunnelkreuzkryotrons. Josephson-Tunnelelemente zeichnen sich durch extrem kurze Schaltzeiten von < 1 ns aus und stellen damit eine echte Alternative zur Halbleitertechnik auf der Grundlage von Silicium und Galliumarsenid für künftige Digitalrechner mit einigen 10^8 Operationen in der Sekunde dar. Theoretisch ließe sich mit Josephson-Tunnelelementen die Zykluszeit eines Rechners auf < 1 ns verkürzen, was einer Taktfrequenz von > 1 000 MHz entspräche. Da die Grenzgeschwindigkeit für Signale die Lichtgeschwindigkeit ist, dürfte ein solcher Rechner nur eine Seitenlänge von < 10 cm haben!
Das Produkt aus Verlustleistung und Schaltzeit ist für Josephson-Tunnelelemente um drei bis vier Größenordnungen niedriger als in der Halbleitertechnik erreichbar, was aus thermischen Gründen Packungsdichten von 10^6 Josephson-Tunnelelementen auf 1 cm² ermöglichen würde. Für ein Tunnelelement liegen die typischen Verlustleistungen zwischen 100 nW und 1 µW, und der Flächenbedarf bei 10 µm². Für die Herstellung können die Fertigungstechnologien der Halbleitertechnik (Foto- und Elektronenstrahllithografie, Bedampfen, Sputtern) übernommen werden. Die Beherrschung der Herstellungsverfahren stellt jedoch hohe Forderungen, da Herstellungs- und Arbeitstemperatur (4,2 K) der Josephson-Tunnelelemente extrem voneinander abweichen, worauf Dünnschichtsysteme empfindlich reagieren. Unbequem ist auch die Notwendigkeit eines ständig vorhandenen Kältesystems. Liegt am Josephson-Übergang eine Spannung U an, die kleiner als die Energielücken der Supraleiter ist, so kommt es zur Anregung eines Wechselstromes der Kreisfrequenz $\omega = q \cdot U/\hbar$ (*Wechselstrom-Jo-*

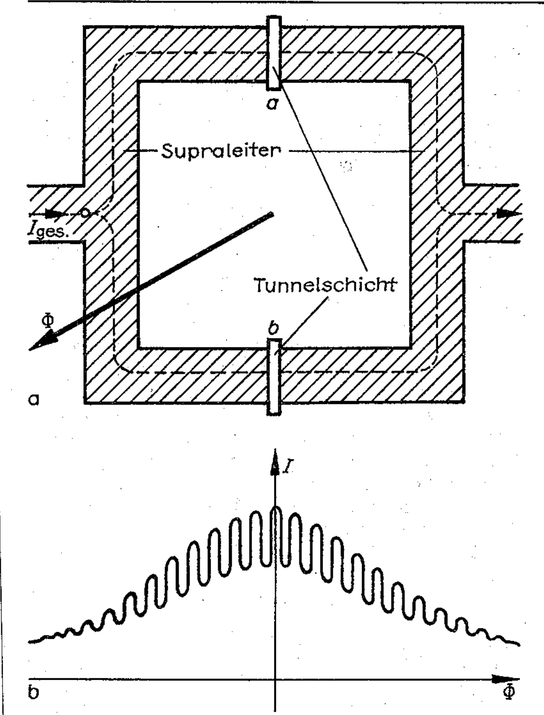


Abb. 8.1.4–6 Supraleitendes Quanten-Interferenz-Bauelement (SQUID), (a) aus zwei Josephson-Tunnelelementen bestehendes SQUID (Schema), (b) Strominterferenzen in Abhängigkeit vom magnetischen Fluß

sephson-Effekt) ($q = 2e$ = Ladung eines Cooper-Paares, $\hbar = h/2\pi$). Für $U = 10\,\mu$V ergibt sich danach eine Frequenz f von $f = \omega/2\pi = 4{,}835$ GHz. Die Genauigkeit, mit der der Zusammenhang zwischen der Kreisfrequenz ω und der der Spannung U meßbar ist, erlaubt eine Neubestimmung von Naturkonstanten, z. B. der Elektronenladung e und des Planckschen Wirkungsquantums h. Ein Josephson-Tunnelelement stellt somit eine kohärente Strahlungsquelle bis zum langwelligen Infrarot dar, die kaum auf andere Art zu verwirklichen ist. Umgekehrt läßt die Einstrahlung einer entsprechenden Frequenz ω in den Josephson-Übergang die Spannung U entstehen.

Josephson-Übergänge reagieren empfindlich auf magnetische Felder. Das wird für eine spezielle Anordnung, das *supraleitende Quanten-Interferenz-Bauelement (SQUID)* genutzt. Den schematischen Aufbau zeigt Abb. 8.1.4–6.

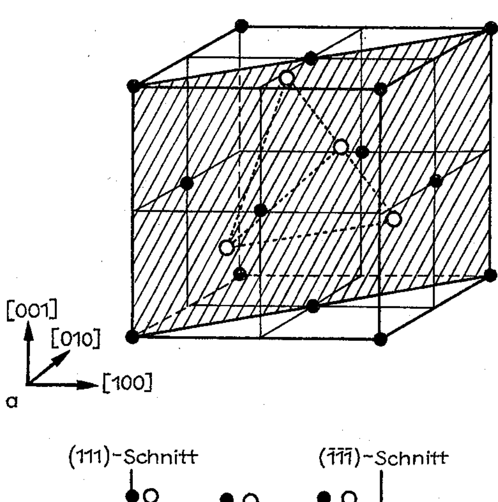
Es kann auf gewisse Analogien mit den Beugungserscheinungen an optischen Gittern verwiesen werden. Während dort Interferenzen von durch zwei Spalte gehenden Amplituden einer Lichtwelle entstehen, interferieren hier Ströme von Cooper-Paaren, die eine starre Phasenkopplung untereinander und damit Kohärenzeigenschaften haben. Die erreichbare Grenzstromstärke I_{Gr} ergibt sich als Funktion des magnetischen Flusses zu $I_{Gr} = 2I_0|\cos e\Phi/\hbar|$. Sie nimmt jedesmal bei ganzzahligen n in $\Phi = n\pi\hbar/e$ ihren Maximalwert an. Der Strom durch das supraleitende Quanteninterferenz-Bauelement ist also immer dann maximal, wenn der magnetische Fluß ein Vielfaches des Flußquants $\Phi_0 = \pi\hbar/e = 2{,}07 \cdot 10^{-15}\,V_s$ beträgt.

Wenn der magnetische Fluß eine Fläche von 1 mm² durchsetzt, ist der Abstand zwischen zwei Strommaxima (vgl. Abb. 8.1.4–6) durch eine magnetische Induktion $B = \Phi$/A der Größe $2 \cdot 10^{-9}$ T gegeben. Da die Ströme bequem auf mindestens Zehntel ihres Maximalwertes gemessen werden können, entspricht das einer Genauigkeit von besser als $2 \cdot 10^{-10}$ T für die magnetische Induktion.

Diese Präzision kommt der der interferometrischen Bestimmung der Lichtwellenlänge gleich. SQUIDs stellen damit hochempfindliche Magnetometer dar, für die sich zahlreiche Anwendungsmöglichkeiten eröffnen. Mit SQUIDs sind z.B. die Magnetfelder meßbar, die von Herz- und Hirnströmen verursacht werden. Die Entdeckung des Josephson-Effektes gehört mit der des Transistoreffektes und des Laserprinzips zu den bedeutendsten physikalischen Leistungen der letzten Jahrzehnte.

8.1.5. Piezoelektrische Bauelemente

8.1.5.1. Piezoelektrischer Effekt. Piezoelektrizität ist eine Kristalleigenschaft. Als (direkten) piezoelektrischen Effekt bezeichnet man das als Folge einer mechanischen Defor-



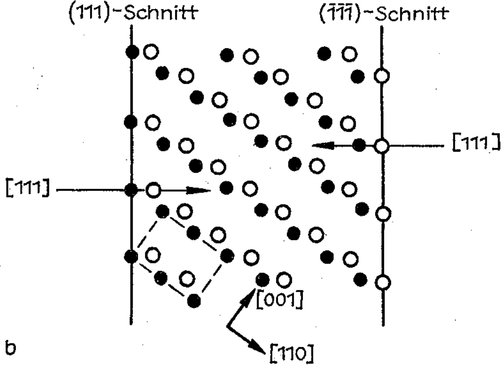


Abb. 8.1.5–1 (a) Elementarzelle der Zinkblendestruktur mit Schnitt in der (1$\bar{1}$0)-Ebene, (b) Schnitt durch die (1$\bar{1}$0)-Ebene; schraffiert: Anteil des Schnitts in (a) mit eingezeichneter polarer Richtung [111] und Gegenrichtung [$\bar{1}\bar{1}\bar{1}$]. ● Zn, ○ S

mation entstehende elektrische Dipolmoment im Kristallinneren und damit eine nach dem Vorzeichen entgegengesetzte elektrische Aufladung einander gegenüberliegender Kristalloberflächen. Piezoelektrizität wird nur bei Substanzen beobachtet, die in Kristallklassen mit polaren Achsen eingeordnet werden können. Bei polaren Achsen sind bestimmte physikalische Eigenschaften von Richtung und Gegenrichtung nicht identisch. Gitterphysikalisch können die elektrischen Ladungen an den Kristalloberflächen durch relative Verrückungen von positiven und negativen Ionen innerhalb des Kristallgitters erklärt werden. Das läßt sich an der Zinkblendestruktur des Zinksulfids zeigen, dessen Bindungskräfte neben dem kovalenten auch einen ionaren Anteil enthalten. In der Elementarzelle des Zinkblendegitters haben die Atomarten Zink und Schwefel in Einheiten der Gitterkonstanten folgende räumliche Koordinaten (Abb. 8.1.5–1a):

Zn: 0, 0, 0 1/2, 1/2, 0 1/2, 0, 1/2 0, 1/2, 1/2
S: 1/4, 1/4, 1/4 3/4, 3/4, 1/4 3/4, 1/4, 3/4
1/4, 3/4, 3/4

Ein Schnitt durch die kubische Elementarzelle, der die beiden senkrechten Würfelkanten vorn links und hinten rechts sowie die zugehörigen Flächendiagonalen der Grund- und Deckfläche enthält, ergibt nach dem Anfügen weiterer Elementarzellen Abb. 8.1.5–1b. Dabei wurde die Richtung der Raumdiagonale [111], die in dem Schnitt enthalten ist, in

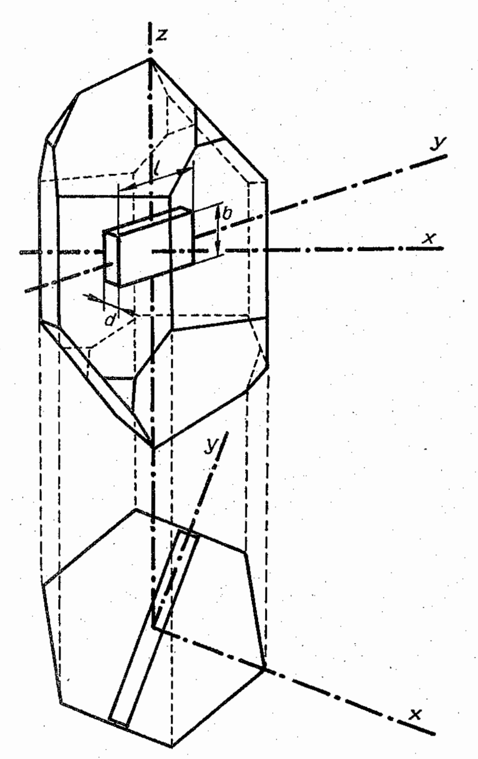


Abb. 8.1.5–2 Quarzeinkristall (Kristallklasse 32) und Orientierung eines Schwingquarzes; z = dreizählige Achse, x = Schwingungsachse bei Dickenschwingungen, y = Schwingungsachse bei Längsschwingungen

die horizontale Richtung gedreht. Es fällt auf, daß in Richtung der Raumdiagonale der Abstand von einem Zink- zu einem Schwefelatom eine Längeneinheit beträgt, während in gleicher Richtung das nächste Zinkatom von dem gerade erreichten Schwefelatom drei Einheiten entfernt ist. In der entgegengesetzten $[\bar{1}\bar{1}\bar{1}]$-Richtung beträgt gerade der Abstand von einem Schwefel- zu einem Zinkatom eine Längseinheit und das nächste Schwefelatom ist von dem eben erreichten Zinkatom drei Längeneinheiten entfernt. Die [111]-Richtung der Zinkblendestruktur ist demnach eine polare Richtung. Bei Anwendung eines Drucks in Richtung einer Raumdiagonale wird der größere Zn-S-Abstand stärker beeinflußt als der kleinere. Entsprechend dem ionaren Bindungsanteil tragen die Atome Überschußladungen verschiedenen Vorzeichens, deren Verschiebung das Auftreten scheinbarer Ladungen an den Kristalloberflächen, also Piezoelektrizität, hervorruft. Bei Druck auf {111}-Flächen lädt sich infolgedessen (111) positiv und $(\bar{1}\bar{1}\bar{1})$ negativ auf.
Der piezoelektrische Effekt wird vor allem beim Quarz und bei piezoelektrischen Keramiken praktisch genutzt.

8.1.5.2. Schwingquarz. In geeigneter Weise aus Quarzeinkristallen herausgeschnittene Plättchen, Scheiben, Stäbchen oder Ringe führen aufgrund des reziproken piezoelektrischen Effektes beim Anlegen eines hochfrequenten elektrischen Wechselfeldes elastische Deformationsschwingungen aus und stellen somit elektromechanische Wandlerbauelemente dar.
Die Anisotropie des piezoelektrischen Effektes und die Nutzung unterschiedlicher Schwingungsmoden ermöglicht das Überstreichen eines Frequenzspektrums von ≈ 1 kHz bis zu mehr als 100 MHz. Allerdings können die hohen Frequenzen nicht mehr mit der Grundschwingung erreicht werden, hierzu ist die Anregung von Oberschwingungen nötig. Eine Quarzscheibe von 5 mm Durchmesser, die bei 10 kHz auf die Grundschwingung angeregt werden soll, darf nur eine Dicke von 0,17 mm haben. Je nach der kristallografischen Orientierung des aus dem Quarzeinkristall (Abb. 8.1.5–2) herausgeschnittenen Stückes und der Elektrodenanordnung sind unterschiedliche Schwingungsarten realisierbar. Die angewandten Schwingungstypen (Abb. 8.1.5–3) und die dazugehörigen Frequenzintervalle sind in Tab. 8.1.5–4 aufgeführt.
Die als Ausgangsmaterial erforderlichen Quarzeinkristalle werden durch Hydrothermal-Synthese hergestellt, einem Verfahren, das der natürlichen Entstehung des Quarzes entspricht und mit einer wässrigen alkalischen Lösung arbeitet, in der sich unter hohem Druck und bei ≈ 400 °C der Quarzbruch auflöst und an einem in dem Druckgefäß hängenden Kristallkeim anlagert. Zur Vorbereitung der Schneidprozedur ist eine sorgfältige kristallografische Orientierungsbestimmung mit einer Genauigkeit von wenigen Winkelminuten auf einem Röntgengonimeter erforderlich.
Neben ihrer Resonanzfrequenz sind Schwingquarze durch ihren Temperaturkoeffizienten der Resonanzfrequenz und durch ihre Güte gekennzeichnet. Die Temperaturabhängigkeit der Frequenz ist i. allg. gering. Bei einer Temperaturschwankung $\Delta T = \pm 20$ °C um die Raumtemperatur ist die relative Frequenzänderung $\Delta f/f$ für die meisten Quarzschnitte von der Größenordnung 10^{-5} und kleiner. Für ausgewählte Schnitte kann eine von der Temperatur unabhängige Frequenz erzielt werden (sogenannter GT-Schnitt).
Die Güte Q, d. h. das Verhältnis der Blindleistung zur umgesetzten Wirkleistung bei der Resonanzfrequenz f_R ist für Schwingquarze im Vergleich zu Schwingkreisen mit Kon-

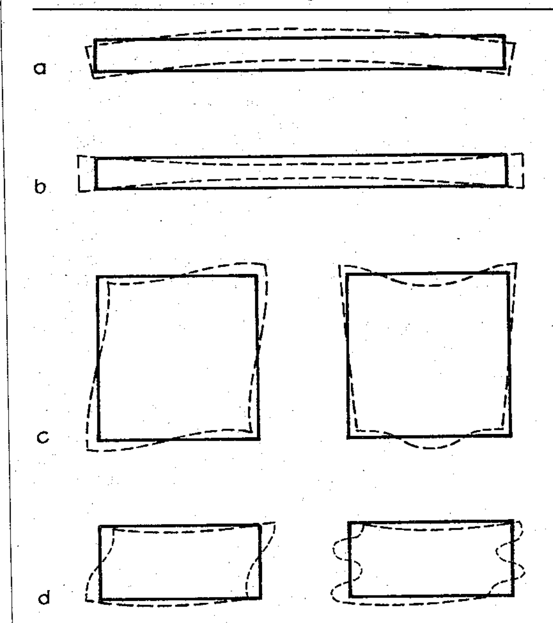


Abb. 8.1.5-3 Schwingungstypen speziell orientierter Quarzplättchen, (a) Längsbiegeschwinger, (b) Längsdehnungsschwinger, (c) Flächenscherschwinger in Grundschwingung und mit Harmonischen, (d) Dickenschwingungen in Grundschwingung und mit dritter Harmonischer

densatoren und Spulen ungewöhnlich hoch und liegt zwischen 10^4 und einigen 10^6.
In seinen elektrischen Eigenschaften kann der Schwingquarz (Schaltsymbol Abb. 8.1.5-5a) durch das in Abb. 8.1.5-5b dargestellte Ersatzschaltbild beschrieben werden. C_1 stellt die elektrostatische Kapazität dar, die sich mit den aufgedampften, aufgesprühten oder aufgepreßten Elektroden und dem Quarz als Dielektrikum ergibt. Die Induktivität L kann als elektrisches Äquivalent der Kristallmasse aufgefaßt werden. Die Kapazität C und der Widerstand R sind die elektrischen Analoga zu mechanischer Nachgiebigkeit und zur mechanischen Reibung bzw. Dämpfung. Sie hängen von Orientierung, Größe und Schwingungstyp des Quarzes ab. Entsprechend dem Ersatzschaltbild hat der Schwingquarz eine Serienresonanzfrequenz f_s (minimaler Scheinwiderstand) und eine Parallelresonanzfrequenz f_p, wobei $f_s < f_p$ und $(f_p - f_s)/f_p \approx 1\,\%$ sind.
Da der kapazitive Blindwiderstand klein gegenüber dem induktiven Blindwiderstand ist, kann für die Güte angegeben werden: $Q = 2\pi f_s L/R$. Als Beispiel sind geometrische, mechanische und elektrische Größen eines Schwingquarzes in Tab. 8.1.5-6 aufgeführt. Die elektrische Belastbarkeit von Schwingquarzen muß so gering gehalten werden, daß die angeregten Schwingungen keine Zerstörung oder Alterung der Anordnung bewirken. Die am Widerstand R auftretende Wirkleistung führt zur Erwärmung des Quarzes und damit zu einer Frequenzänderung. Schwingquarze sind in Miniaturröhrenkolben bzw. für höhere Frequenzen in Metallhaltern eingebaut.
Schwingquarze dienen hauptsächlich der Erzeugung von Schwingungen mit hoher Frequenzkonstanz. Eine Quarzoszillator-Schaltung zeigt Abb. 8.1.5-7. Ihre Schwingungsfre-

Tab. 8.1.5-4 Typen von Schwingquarzen

Typ	Frequenz-intervall	wirksame elektrische Feldstärke (vgl. Abb. 8.1.5-2)
Längsbiegungsschwinger (vgl. Abb. 8.1.5-3a)	1...50 kHz	$\pm E_x$
Längsdehnungsschwinger (vgl. Abb. 8.1.5-3b)	50...150 kHz	E_x
Flächenscherschwinger (vgl. Abb. 8.1.5-3c)	150...500 kHz	E_y
Dickenscherschwinger (vgl. Abb. 8.1.5-3d)	0,3...100 MHz	E_y
Dickendehnungsschwinger (Anregung der Grundschwingung)	0,2...50 MHz	E_x

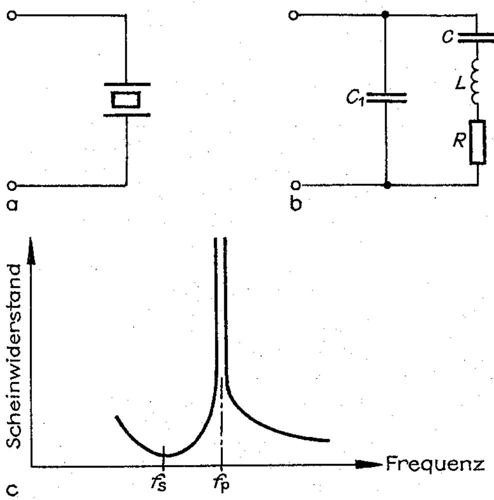


Abb. 8.1.5-5 Schwingquarz, (a) Symbol, (b) Ersatzschaltbild, (c) Verlauf des Scheinwiderstands in der Nähe der Serien- und Parallelresonanzfrequenz f_s und f_p

Tab. 8.1.5-6 Geometrische, mechanische und elektrische Größen eines Schwingquarzes

Geometrie
Länge 2,75 cm
Breite 3,33 cm
Dicke 0,636 cm

Resonanzfrequenzen (Dickenscherschwinger)
Serienresonanzfrequenz $f_s = 427{,}4$ kHz
Parallelresonanzfrequenz $f_p = 430{,}1$ kHz

elektrische Größen
Induktivität $L = 3{,}3$ H
Serienkapazität $C = 0{,}042$ pF
Parallelkapazität $C_1 = 5{,}8$ pF
Güte $Q = 25\,000$

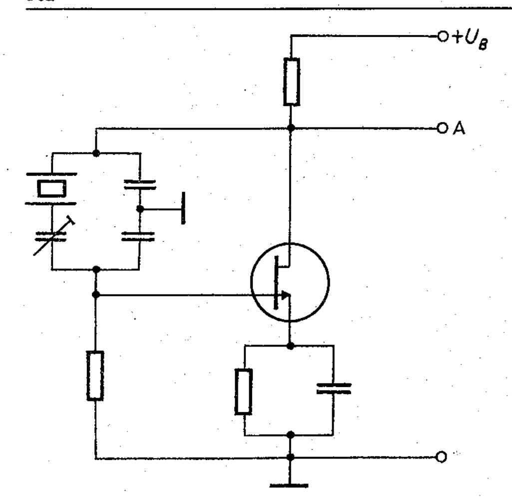


Abb. 8.1.5–7 Schaltung eines Quarzoszillators mit Feldeffekttransistor (Pierce-Oszillator)

quenz liegt zwischen f_s und f_p und erfolgt auf der Mode der höchsten Aktivität.

Bei Verwendung ausgesuchter Quarze, Thermostatierung des gesamten Oszillators und Stabilisierung der Stromversorgung ist eine Frequenzkonstanz von $\Delta f/f \approx 10^{-9}$ erreichbar. Quarzoszillatoren eignen sich damit vorzüglich als Frequenz- bzw. Zeitnormale für Meß- und Eichzwecke, z. B. zur Erzeugung einer konstanten Trägerfrequenz bei Rundfunksendern. Sie sind auch wesentlicher Bestandteil von Quarzuhren, deren Ganggenauigkeit durch die erreichbare Frequenzkonstanz bestimmt wird. Quarzuhren sind vom Prinzip her Impulszähler. Durch Frequenzteiler (vgl. 8.2.4.4.) wird aus der Resonanzfrequenz des Schwingquarzes der Sekundenimpuls erzeugt. Mit der Entwicklung der Mikroelektronik sind mechanische Armbanduhren weitgehend durch elektronische Armbanduhren verdrängt worden. Digitaluhren mit Schwingquarz und Flüssigkristallanzeige (vgl. 8.1.6.9.) bestehen nur noch aus fünf zusammenzufügenden Einzelteilen (Schwingquarz, verbreitet sind Resonanzfrequenzen von 32,768 kHz = 2^{15} Hz und 4,194304 MHz = 2^{22} Hz; integrierter Schaltkreis; Flüssigkristallanzeige; Batterie; Gehäuse) und haben keine mechanisch bewegten Teile. Demgegenüber werden mechanische Armbanduhren aus mehr als 100 Einzelteilen zusammengefügt, zu deren Herstellung und Montage über 1000 Arbeitsgänge nötig sind. Die unbequemere Ablesung mehrerer Ziffern gegenüber zwei Zeigerstellungen kann durch Flüssigkristall-Zeigertableaus überwunden werden. Elektronische Armbanduhren haben in der Massenproduktion eine Gangunsicherheit von wenigen Minuten je Jahr.

8.1.5.3. Piezokeramik. Unter den piezokeramischen Werkstoffen sind Mischkristalle aus Bleititanat ($PbTiO_3$) und Bleizirkonat ($PbZrO_3$), die gleichzeitig ferroelektrisch sind, am verbreitetsten. Beide Substanzen bilden eine vollständige Mischkristallreihe. Durch Zuschlagstoffe in Form von Metalloxiden wird eine Reihe von Eigenschaften wie Dielektrizitätskonstante, Curie-Temperatur, Güte, Witterungsbeständigkeit und Temperaturabhängigkeit beeinflußt. Da piezokeramische Werkstoffe polykristallin sind, können sie in vielfältigen Formen und mit geeigneten Texturen hergestellt werden. Zur Charakterisierung sind nur die piezoelektrischen Koeffizienten d_{33} und d_{31} nötig, die für Piezokeramik um etwa zwei Zehnerpotenzen höher liegen, als für Quarz. Piezokeramik wird als Werkstoff in der Filtertechnik genutzt, dient aber auch zur Herstellung von Zündelementen für Feuerzeuge und Gasanzünder und für elektroakustische Wandler (Mikrofone, Tonabnehmer). Weitere Einsatzgebiete sind Ultraschalldiagnose, -therapie und das Echolot. Der unter dem Einfluß äußerer mechanischer Spannungen hervorgerufene zusätzliche Anteil am spezifischen elektrischen Widerstand heißt Piezowiderstand. Er wird durch einen Tensor vierter Stufe angegeben. Ein Anwendungsbeispiel ist der Halbleiterdehnungsmeßstreifen (vgl. 8.1.1.7.).

8.1.5.4. Der akustoelektrische Effekt und seine Anwen-

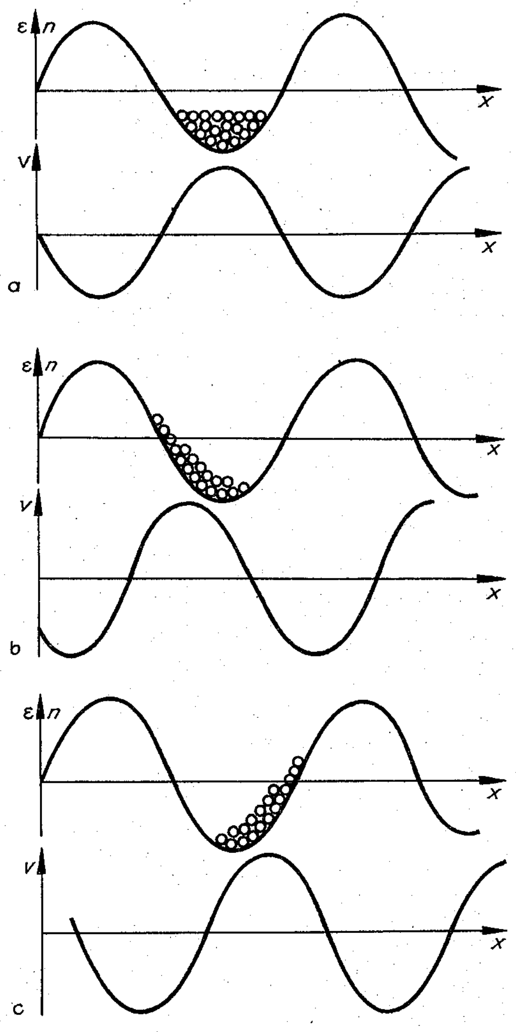


Abb. 8.1.5-8 Periodische Gitterdeformation $\varepsilon(x)$ durch die Schallwelle, potentielle Energie $V(x)$ der Ladungsträger, Ladungsträgerkonzentration n, (a) Geschwindigkeit der Schallwelle v_s = Driftgeschwindigkeit der Ladungsträger v_D, (b) $v_s > v_D$, (c) $v_s < v_D$

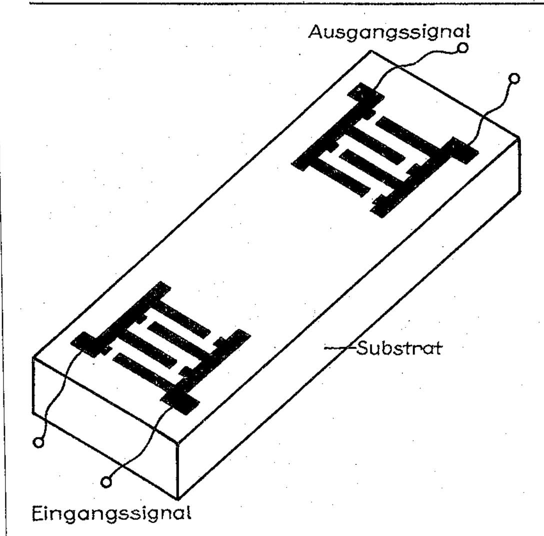


Abb. 8.1.5–10 Akustoelektronischer Wandler mit piezoelektrischem Substrat (Interdigitalwandler)

dung. Der akustoelektrische Effekt hat die gegenseitige Beeinflussung von Schallwellen mit den für die Transportphänomene verantwortlichen und im Gitterpotential quasifrei beweglichen Ladungsträgern in Halbleitern zum Inhalt. Da Schallwellen im Teilchenbild als Phononen, Schallquanten, darstellbar sind, enthält der akustoelektrische Effekt eine spezielle Seite der Elektron-Phonon-Wechselwirkung. Es wird vorwiegend der Einfluß eines „Schallstrahles", d. h. einer Vielzahl gerichteter Phononen betrachtet. Eine einem Halbleiter zugeführte Ultraschalleistung P hat eine akustoelektrische Stromdichte I_{ae} zur Folge, die von der Ladungsträgerbeweglichkeit μ des Halbleiters und der Dämpfungskonstante α_e abhängt, $I_{ae} = 2\,\mu\alpha_e P$.

Da die Dämpfung des Schalls der Übernahme von Schallenergie durch die Ladungsträger entspricht, geht bei Gültigkeit von Energie- und Impulssatz auch der Phononenimpuls auf die Ladungsträger über, wobei die Dämpfung der Ladungsträgerkonzentration proportional ist. Diese Wechselwirkung ist besonders für Halbleiter, die gleichzeitig Piezoelektrika sind, in Richtung polarer Achsen (vgl. 8.1.5.1.) stark. Die Schallwelle bewirkt eine lokale Verschiebung der Gitterbausteine und ruft damit eine Deformationswelle im piezoelektrischen Halbleiter hervor. Für die quasifreien Ladungsträger ist das eine Welle der potentiellen Energie und die Ladungsträger reichern sich in den Minima der Welle an. Für die zu erwartenden Erscheinungen sind die Fortpflanzungsgeschwindigkeit der Schallwelle v_s und die Driftgeschwindigkeit der Elektronen v_D zu vergleichen (Abb. 8.1.5–8). Für $v_S = v_D$ wird sich den Ladungsträgern eine durch die Potentialwelle gegebene örtliche Konzentrationsabhängigkeit aufprägen. Da für diesen Fall Elektronen und Phononen gleiche Geschwindigkeit haben, finden keine Stöße statt und die Elektronen, die dann Mach-1-Geschwindigkeit haben, tauschen keine Energie aus. Stimmen v_D und v_S nicht überein, so findet ein Energieaustausch statt. Schneller driftende Ladungsträger übertragen Energie auf die Schallwelle, verstärken sie, während langsamer driftende Elektronen Energie vom Schallfeld aufnehmen. Bei genauerer Betrachtung sind neben der Anisotropie des Materials eine Vielzahl von Vorgängen, z. B. die frequenzabhängige dielektrische Relaxation der Substanz, die Ladungsträgerdiffusion, das den Ladungsträgertransport beeinflussende Störstellenspektrum des Halbleiters und die durch die Schallwellen angeregten Gitterschwingungen, zu berücksichtigen. Das in der Forschung intensiv bearbeitete Gebiet stellt neuartige Anwendungen bereit und entwickelte sich zur disziplinellen Richtung der Akustoelektronik als Zweig der Festkörperphysik.

Eine besondere Rolle in der Anwendung akustoelektronischer Erscheinung spielen *Oberflächenwellen* (Abb. 8.1.5–9). Da an der Oberfläche eines Festkörpers die Bindung der Bausteine nur durch den das Festkörperinnere darstellenden Halbraum erfolgt, erleiden Oberflächenwellen auch eine geringere Dämpfung als Volumenwellen. Das ist von Erdbebenwellen her bekannt. Mit Hilfe eines speziellen piezoelektrischen Wandlers können Oberflächenwellen mit einer Eindringtiefe von wenigen Mikrometern verlustarm angeregt werden. Der Wandler besteht aus einer piezoelektrischen Substratplatte aus Quarz oder Lithiumniobat ($LiNbO_3$), an deren Enden zwei Paare kammartig ineinandergreifender Metallelektroden mit einer Schichtdicke von wenigen Zehntel Mikrometern aufgebracht werden (vgl. Abb. 8.1.5–10). Der Abstand benachbarter „Finger" ist auf eine halbe Wellenlänge der anzuregenden Oberflächenwelle eingestellt. Ein an ein Elektrodenpaar angelegtes periodisches Signal erzeugt in der piezoelektrischen Substanz ein elektrisches Feld, das Deformationswellen an der Ober-

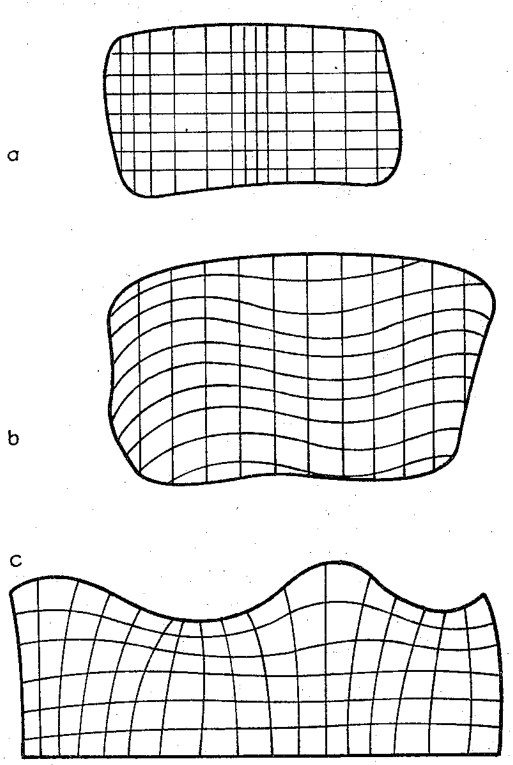


Abb. 8.1.5–9 Elastische Wellen in Festkörpern, (a) longitudinale und (b) transversale Volumenwelle, (c) Oberflächenwelle

fläche zur Folge hat, die sich in beiden Richtungen senkrecht zu den Fingern der Kammstruktur ausbreiten. Die sich über die Substratplatte zum anderen Elektrodenpaar bewegende Welle kann dort als elektrisches Signal wieder abgenommen werden. Die Fortpflanzungsgeschwindigkeit des Signals ist dabei die Schallgeschwindigkeit im piezoelektrischen Medium. Da diese nur das 10^{-5}fache der Lichtgeschwindigkeit beträgt, ist ein solches Wandlerbauelement gleichzeitig eine Verzögerungsleitung. *Verzögerungsleitungen* (Laufzeitketten) werden zur Verzögerung durchlaufender Signale verwendet und z. B. bei Koinzidenz-Schaltungen von Detektorimpulsen bei Kernreaktionen, zur Erzeugung höchstfrequenter Schwingungen mit Laufzeitröhren oder im Farbfernsehverfahren SECAM zur Verzögerung eines Zeilensignals um 64 µs benötigt. Während größere Signalverzögerungen auf rein elektrischem Wege beträchtlichen Aufwand erfordern, stellt der piezoelektrische Wandler der Abb. 8.1.5–10 eine Verzögerungsleitung auf kleinstem Raum dar. Mit piezoelektrischen Wandlern nach Abb. 8.1.5-10, die nach den in der Mikroelektronik üblichen fotolithografischen Verfahren hergestellt werden, können vorteilhaft Oberflächenwellenfilter realisiert werden, die als Bandpässe arbeiten. Als *Bandpaß* bezeichnet man ein Übertragungsglied, das nur für ein bestimmtes Frequenzintervall durchlässig ist. Das Frequenzintervall, für das die Anordnung durchlässig ist, wird dabei durch die geometrischen Dimensionen der Kammstruktur bestimmt. Vorteilhaft sind die kleinen äußeren Abmessungen für Frequenzen zwischen 10 und 1000 MHz, die über die Lebensdauer des Filters stabil bleibenden Selektionseigenschaften und der wegfallende Abgleich des Filters. Neben Lithiumniobat ($LiNbO_3$) und Quarz werden als Substrate für Oberflächenwellenfilter häufig auch Lithiumtantalat ($LiTaO_3$), Bismutgermaniumoxid ($Bi_{12}GeO_{20}$) und Piezokeramik verwendet.

Ebenso ist eine Anwendung in der Lichtleittechnik denkbar. Durch Anregung stehender Schallwellen könnte in den Lichtweg ein durch elektrische Signale erzeugbares Beugungsgitter eingefügt werden. Dem entspricht eine elektrisch zu betätigende Weiche für das Licht (vgl. 8.1.6.10.).

8.1.6. Optoelektronische Bauelemente

Die Optoelektronik nutzt als Teilgebiet der Elektronik die physikalischen Prinzipien der Wechselwirkung zwischen elektrischen Ladungsträgern und Photonen des sichtbaren Teils einschließlich des infraroten und ultravioletten Bereiches des elektromagnetischen Spektrums zur Informationsgewinnung, -übertragung, -verarbeitung und -speicherung aus. Die Optoelektronik entwickelt sich rasch und findet Zugang zu zahlreichen technischen Anwendungen. Als Bestandteil von Elektronik und Halbleitertechnik konnten in der Optoelektronik gewonnene Erkenntnisse wesentliche Beiträge zur Entwicklung des gesamten Gebiets liefern.

Lumineszenz- und Laserdioden gewährleisten die modulierbare Strahlenemission z. B. über die Intensität der elektromagnetischen Strahlung, Fotodetektoren dienen der Rückverwandlung von der optischen in die elektrische Signalform. Die optische Strahlung mit ihrer hohen Trägerfrequenz der Größenordnung 10^{14} Hz (Rundfunk, Fernsehen, Richtfunk liegen im elektromagnetischen Spektrum zwischen 10^5 Hz und $\approx 10^{10}$ Hz) ermöglicht für die Signalfrequenzen gegenüber denen in der konventionellen Nachrichtentechnik eine um mehrere Größenordnungen anwachsende verfügbare Bandbreite. Die Signalübertragung über Glasfasern hat, neben der Ersparnis metallischer Leiter, den Vorteil der Störsicherheit gegenüber äußeren elektromagnetischen Feldern.

8.1.6.1. Fotowiderstände bestehen aus sperrschichtfrei kontaktierten, homogenem Halbleitermaterial. Sie ändern ihre elektrische Leitfähigkeit $\sigma = e(\mu_n n + \mu_p p)$ (e = Elementarladung, n, p = Elektronen- bzw. Defektelektronenkonzentration, μ_n, μ_p = Beweglichkeit der Elektronen bzw. der Defektelektronen) in Abhängigkeit von der Beleuchtungsstärke. Für einkristalline Fotohalbleiter ist die durch den inneren Fotoeffekt bewirkte Erhöhung der Ladungsträgerkonzentration der entscheidende Mechanismus. Für polykristallines Material tritt als zusätzlicher Mechanismus die an den Korngrenzen wirksame Erhöhung der Ladungsträgerbeweglichkeit (Quotient aus Driftgeschwindigkeit und elektrischer Feldstärke) hinzu. Als ein- bzw. polykristalline Fotowiderstände sind sowohl eigenleitende als auch bewußt dotierte Halbleitersubstanzen verwendbar.

Für die eigenleitenden Materialien erfolgt die Bereitstellung von Ladungsträgern für die Transportvorgänge durch die Grundgitterabsorption. Die absorbierten Photonen übergeben ihre Energie an Ladungsträger im Valenzband, die mit diesem Energiegewinn in das Leitungsband gehoben werden und quasifrei beweglich sind. Fotowiderstände aus Germanium und Silicium sind dotiert, d.h. bewußt mit Störstellen versehen. Bei Photonenabsorption gehen entweder Ladungsträger aus dem Valenzband in Störstellen und lassen im Valenzband bewegliche Löcher zurück, oder Ladungsträger werden von Störstellen in das Leitungsband gebracht und nehmen dort am Transport teil. Fotowiderstände werden vorwiegend aus $A_{II}B_{VI}$-Verbindungen hergestellt. Häufig angewandt werden die Sulfide, Selenide und Telluride des Cadmiums. Daneben werden die $A_{III}B_V$-Verbindungen Indiumarsenid (InAs) und Indiumantimonid (InSb) und die $A_{IV}B_{VI}$-Verbindungen Bleisulfid (PbS), Bleiselenid (PbSe) und Bleitellurid (PbTe), die Bleichalkogenide, benutzt. Fotowiderstände aus Verbindungshalbleitern arbeiten mit Grundgitterabsorption.

Das Photon trägt die Energie $W = h\nu$ (ν = Frequenz, $h = 6{,}626 \cdot 10^{-34}\ \mathrm{Ws^2}$, Plancksche Konstante). Für seine Absorption kann demnach eine minimale Frequenz des Lichtes bzw. eine maximale Wellenlänge (Grenzwellenlänge) angegeben werden, für die eine Anregung von Ladungsträgern gerade noch möglich ist. Dazu muß die Energie der Photonen bei Grundgitterabsorption die Energie des Abstands zwischen oberer Kante des Valenzbands und unterer Kante des Leitungsbands haben: $\lambda_{grenz} = ch/W_L$ (W_L = Band-

Tab. 8.1.6-1 Daten von Fotowiderständen

Substanz	Wellenlänge der maximalen Empfindlichkeit in µm	Ansprechzeit in µs
Cadmiumsulfid (CdS)	0,5	100...50000
Cadmiumselenid (CdSe)	0,7	100...50000
Bleisulfid (PbS)	2,1	100...250
Bleiselenid (PbSe)	3,4	50
Indiumantimonid (InSb)	6	0,1
Germanium : Kupfer (Ge : Cu)	20 (bei 10 K)	0,01...1

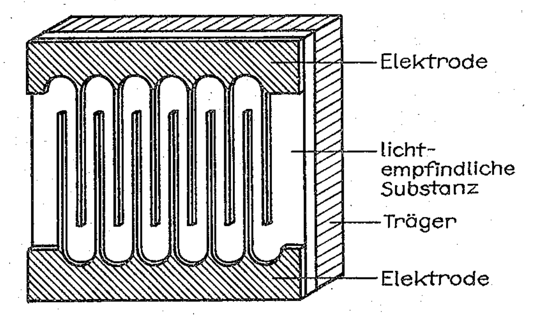


Abb. 8.1.6–2 Schema eines Fotowiderstandes

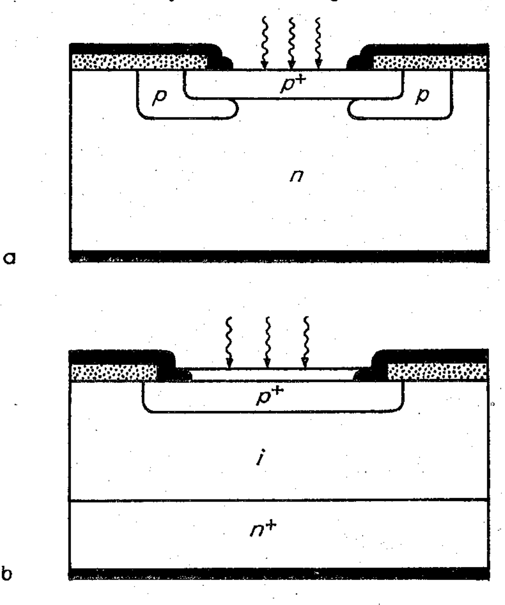
abstand, c = Lichtgeschwindigkeit). Für Cadmiumsulfid (CdS) bzw. Cadmiumselenid (CdSe) mit Bandabständen von 2,4 eV bzw. 1,7 eV liegen danach die Grenzwellenlängen bei 520 nm bzw. 740 nm (vgl. Tab. 8.1.6–1). Fotowiderstände weisen eine relativ große Ansprechzeit (Zeitdauer zwischen Beginn der Lichteinstrahlung und Ausbildung eines konstanten Fotostroms) auf, die bis an eine Zehntelsekunde heranreichen kann. Im Infrarot empfindliche Fotowiderstände werden häufig im gekühlten Zustand betrieben, da der Dunkelstrom durch thermische Anregung bei Raumtemperatur wegen der geringen Bandlücke schon merklich ist. Dunkel- und Hellwiderstand können sich z. B. bei CdS-Fotowiderständen um einen Faktor 10^5 unterscheiden.

Die Elektroden von Fotowiderständen sind meist kammartig ineinandergreifend (Abb. 8.1.6–2) ausgeführt. Fotowiderstände werden zur elektrischen Belichtungsmessung, für Dämmerungsschalter oder Lichtschranken verwendet. Für Anwendungsbereiche, die eine kürzere Ansprechzeit fordern, werden sie durch Fotodioden verdrängt.

8.1.6.2. Fotodioden. In einer Diode, deren wesentlicher Bestandteil ein *pn*-Übergang ist, hängt der Sperrstrom von der Konzentration der Minoritätsträger, also der Elektronen im *p*-Gebiet und der Defektelektronen im *n*-Gebiet, ab. Beim Gebrauch einer Diode als Schalter oder zum Zweck der Gleichrichtung ist ein vernachlässigbar kleiner Sperrstrom ständig erwünscht. Für bestimmte Anwendungen des *pn*-Übergangs wird aber auch die Steuerung des Sperrstromes ausgenutzt. Beispielsweise läßt sich die Beeinflussung des Sperrstromes einer Diode für eine wichtige Energieformwandlung einsetzen. Werden nämlich im Bereich des Übergangsgebietes durch Photonenabsorption Elektronen-Defektelektronen-Paare erzeugt, so erhält die Sperrstromdichte I_{sp} einen zusätzlichen Beitrag. Aufgrund des elektrischen Feldes im Übergangsgebiet werden die durch Lichteinstrahlung entstandenen Ladungsträgerpaare getrennt und nach entgegengesetzten Seiten abgeführt, wodurch der Zusatz I_G zur Sperrstromdichte entsteht. Es ist $I_G = eG(L_n + L_p)$. G beschreibt die Erzeugungsrate von Elektronen-Defektelektronen-Paaren infolge der Photonenabsorption je Zeit- und Volumeneinheit L_n, L_p stellen die Diffusionslängen für Elektronen bzw. Defektelektronen dar. Das sind die mittleren Entfernungen, die die entsprechenden Nichtgleichgewichtsladungsträger durch Diffusion während ihrer „Lebensdauer" t, d. h. bis zu ihrer Rekombination, zurücklegen.

Der lichtempfindliche *pn*-Übergang liegt parallel zur Oberfläche in einer Tiefe von wenigen Mikrometern (Abb. 8.1.6–3a). Das Licht fällt senkrecht auf den *pn*-Übergang. Der mit wachsender Lichteinstrahlung ansteigende Sperrstrom erzeugt an einem mit der Fotodiode in Reihe liegenden Widerstand einen der Strahlungsintensität proportionalen Spannungsabfall. Ein Schutzring am Rand des *pn*-Übergangs verhindert bei hohen Feldstärken einen Durchbruch.

Fotodioden für den sichtbaren Spektralbereich werden aus Silicium und Germanium hergestellt. Siliciumfotodioden sind im Wellenlängenbereich von 0,4 bis 1,1 µm mit einem Maximum bei ≈ 0,9 µm empfindlich. Germaniumfotodioden sprechen zwischen 0,6 und 1,65 µm mit einem Maximum bei ≈ 1,4 µm auf Lichteinstrahlung an. Fotodioden verdrängen zunehmend die Fotovervielfacher, die einen komplizierten mechanischen Aufbau haben und verglichen mit den Fotodioden beträchtliche Volumina einnehmen. Da die Sperrschichtkapazität des *pn*-Übergangs ebenfalls von der Belichtung abhängt, bieten Fotodioden die Möglichkeit, Änderungen der Lichtintensität auch in Frequenzänderungen zu überführen. Von grundsätzlicher Bedeutung sind die Fotodioden jedoch als Strahlungsdetektoren in der



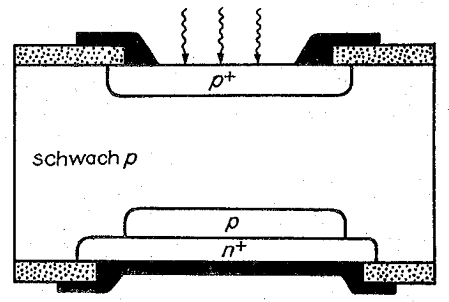


Abb. 8.1.6–3 Fotodioden (schematisch), (a) Schnitt durch eine Silicium-*pn*-Fotodiode mit Schutzring, (b) Schnitt durch eine *pin*-Fotodiode mit reflexionsverringerndem Belag, (c) Lawinenfotodiode

optischen Nachrichtenübertragung über Lichtleiter. Für die Spektralbereiche, bei denen Lichtleiter minimale Dämpfungen aufweisen und für die gleichzeitig effektive Strahlungsemitter in Form von Halbleiter- und anderen Festkörperlasern und Luminenszenzdioden zur Verfügung stehen, z.B. bei 0,85 µm und bei 1,06 µm, eignen sich Strahlungsdetektoren aus Silicium und ternären Mischkristallen von $A_{III}B_V$-Verbindungshalbleitern (z. B. $GaAs_xSb_{1-x}$). Dem Nachteil der schwierigeren Herstellungstechnologie steht bei diesen Mischkristallen der Vorteil gegenüber, die Wellenlänge der maximalen spektralen Empfindlichkeit durch den Mengenanteil der Mischungskomponenten zu steuern.

Für die Informationsübertragung in Lichtleitungssystemen haben sich zwei spezielle Formen von Fotodioden als geeignet erwiesen. Bei der *pin-Fotodiode* (Abb. 8.1.6–3b) ist zwischen das *p*- und *n*-Gebiet des Übergangs eine hochohmige eigenleitende (intrinsic) Schicht eingebaut. Dadurch vergrößert sich einerseits das Volumen, in dem Ladungsträgerpaare erzeugt und abgesaugt werden können, d.h. es wächst die Quantenausbeute. Andererseits verringert sich der Sättigungssperrstrom. Die Dicke der eigentlichen Schicht bestimmt bei gegebenem Absorptionskoeffizienten, wie groß die Wahrscheinlichkeit dafür ist, daß jedes Photon ein Ladungsträgerpaar erzeugt. Diese Wahrscheinlichkeit wächst mit wachsender Dicke der *i*-Schicht. Dann verzögert sich allerdings wegen der größeren Ladungslaufzeiten das Reagieren des Stroms auf Intensitätsänderungen. Je nach Einsatz liegen die Dicken zwischen wenigen Mikrometern und 1 mm. Es gibt auch *pin*-Fotodioden, bei denen das Licht seitlich, parallel zu den Grenzflächen des Übergangs, einfällt.

Bevorzugt als Empfänger in optischen Übertragungssystemen wird eine Fotodiodenform, die einen inneren Verstärkungseffekt hat und daher gegenüber den bisherigen Formen eine erhöhte Empfindlichkeit aufweist. In diesen *Lawinen-Fotodioden (Avalanche-Fotodioden)* werden die durch Lichteinstrahlung erzeugten Elektron-Defektelektron-Paare einer derartig hohen elektrischen Feldstärke ausgesetzt, daß die gewonnene kinetische Energie für die Elektronen ausreicht, um über Stoßionisation weitere freie Ladungsträger zu erzeugen. Die Anzahl der für den Strom zur Verfügung stehenden Ladungsträger erhöht sich bei hinreichend hoher äußerer Spannung lawinenartig. Verstärkungsfaktoren für den Fotostrom liegen zwischen 10^2 (bei Germaniumlawinendioden) und 10^6 (Siliciumlawinendioden). In der Anwendung nachteilig sind die für Halbleiterschaltungen ungewöhnlich hohen Spannungen in der Nähe des Diodendurchbruchs von oft mehreren hundert Volt.

Für Fotodioden, die im infraroten Spektralbereich bis zu ≈ 30 µm empfindlich sind, stehen Halbleitermaterialien mit kleineren Bandabständen bis zu praktisch verschwindendem Bandabstand zur Verfügung. Dafür eignen sich $A_{IV}B_{VI}$- und $A_{II}B_{VI}$-Verbindungen, z.B. PbS, $Pb_{1-x}Sn_xTe$, $Hg_{1-x}Cd_xTe$. Die Mischkristalle haben wieder den Vorteil, daß der Bandabstand über den Komponentenanteil einstellbar ist.

Wenn sich bei einem schichtförmigen, innigen Kontakt zwischen einem Metall und einem Halbleiter eine an Ladungsträgern verarmte Randschicht herausbildet, erhält man ein Bauelement mit Gleichrichtereigenschaften, die *Schottky-Diode.* Wenn es sich um einen *n*-Halbleiter handelt, ist dafür Voraussetzung, daß die Austrittsarbeit (Energie, die man einem auf der Fermi-Energie befindlichem Elektron mitteilen muß, damit es den Festkörper verlassen kann und danach die kinetische Energie null hat) des Metalles W_M größer ist, als die Austrittsarbeit des Halbleiters W_H. Zusätzlich muß der Einfluß von Oberflächenzuständen, die durch die Grenzfläche bedingt sind, berücksichtigt werden. Diese Oberflächenzustände verhalten sich wie Störstellen und schränken die Bedeutung der Austrittsarbeitsdifferenz für das Auftreten einer Verarmungsrandschicht ein. Das elektrische Verhalten eines solchen Metall-Halbleiter-Übergangs wird durch auffallende Lichtstrahlung beeinflußt, wenn die Metallelektrode hinreichend transparent ($d \approx 10$ nm) ist. Bauelemente dieser Art heißen *Schottky-Fotodioden.* Ist die Photonenenergie $h\nu$ größer als der Bandabstand W_L, so entstehen im Halbleiter Elektron-Defektelektron-Paare. Spannt man den Metall-Halbleiter-Übergang mit einer sehr hohen Sperrspannung vor, so läßt sich durch Stoßionisation ein innerer Verstärkungseffekt wie bei der Lawinenfotodiode erzielen.

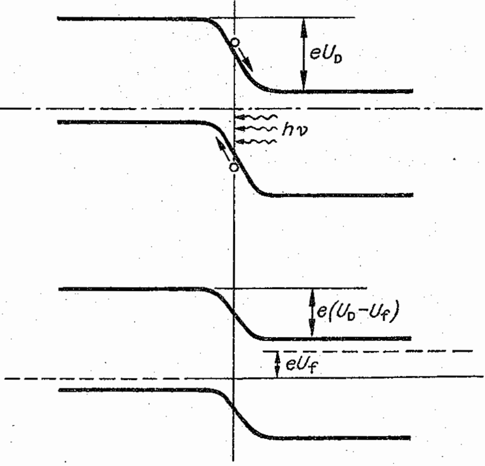


Abb. 8.1.6–4 Entstehung einer Fotospannung U_f am *pn*-Übergang

8.1.6.3. Solarzellen sind Halbleiterfotoelemente, bei denen an einem *pn*-Übergang unter dem Einfluß einfallenden Lichts, dessen Photonenenergie größer ist als der Bandabstand, eine Fotospannung entsteht. Vor der Belichtung sind die Stromdichten der Elektronen bzw. Defektelektronen aus den einander gegenüberliegenden Gebieten gleich groß. Bei Lichteinstrahlung werden Elektron-Defektelektron-Paare erzeugt, die im Raumladungsgebiet des *pn*-Übergangs aufgrund des vorhandenen inneren elektrischen Feldes getrennt wurden. Diese Ladungsträger verhalten sich wie Minoritätsträger und addieren sich zum Sättigungssperrstrom. Dadurch wird die am *pn*-Übergang liegende Potentialbarriere um eU_f erniedrigt (Abb. 8.1.6–4). Diese Erniedrigung tritt gegenüber dem *pn*-Übergang im thermischen Gleichgewicht (vgl. Abb. 8.1.3–1) als Urspannung am *pn*-Übergang, bzw. als Fotostrom I_f auf. Der beleuchtete *pn*-Übergang wirkt als Spannungsquelle. Seine Strom-Spannungs-Kennlinie folgt aus der Gleichung des *pn*-Übergangs (vgl. 8.1.3.1.)
$I = I_{sp}\left\{\exp\left(\frac{eU}{kT}\right) - 1\right\} - I_G$. I_G ist dabei die durch die erzeugten Elektron-Defektelektron-Paare gelieferte Stromdichte. Für I_G gilt die Beziehung $I_G = eG(L_n + L_p)$ (vgl. 8.1.6.2.). Der Fotostrom I_G ist der Kurzschlußstrom, der bei $U = 0$ fließt. Demzufolge ergibt sich bei verschwindendem

Gesamtstrom $I = 0$ eine Leerlaufspannung, die Fotospannung $U_f = \frac{kT}{e} \ln\left(\frac{I_G}{I_{sp}} + 1\right)$.

Der Wirkungsgrad der Energiewandlung von Solarzellen wird als das Verhältnis der entnehmbaren elektrischen Leistung je Quadratzentimeter zur Solarkonstanten (0,1395 mW cm^{-2}) definiert. Mit *p*-Silicium des Störstellengehalts von $N_A = 1{,}3 \cdot 10^{16}$ cm^{-3} und einem spezifischen Widerstand von 1 Ω cm werden Wirkungsgrade zwischen 12 und 14 % erzielt. Es ergeben sich Leerlaufspannungen $U_f = 0{,}6$ V und Kurzschlußstromdichten $I_G = 35$ mA/cm^2. Für irdische und extraterrestrische Energiewandlung in großen Maßstäben sind Solarzellen ökonomisch sehr aufwendig. Eine Kostenreduzierung könnte durch die Verwendung polykristallinen Siliciums oder auch von Cadmiumsulfid (CdS) als Basismaterial zustande kommen.

Die Entstehung von Fotospannungen ist nicht auf *pn*-Übergänge beschränkt. Wenn auf andere Weise die zur Ladungstrennung erforderlichen Felder zustande kommen, können durch Photonenabsorption erzeugte Elektron-Defektelektron-Paare eine Fotospannung hervorrufen. So kommt es auch in homogenen Halbleitern aufgrund des *Dember-Effektes* zu einer Fotospannung: Werden in der Oberflächenschicht eines Halbleiters durch Lichteinstrahlung Elektron-Defektelektron-Paare erzeugt, so diffundieren die Ladungsträger infolge des erzeugten Konzentrationsgradienten ins Innere und auf die gegenüberliegende Oberfläche zu. Die Ladungsträgerart mit der größeren Beweglichkeit diffundiert schneller, die anderen langsamer. Die Folge ist eine Raumladung mit einem ihr innewohnenden elektrischen Feld, das sich als Fotospannung zwischen der belichteten und der unbelichteten Seite eines Halbleiterplättchens nachweisen läßt. Bringt man einen Halbleiterquader, dessen Vorderfläche belichtet wird, in ein Magnetfeld parallel zum Lichtstrom, dann trennt die Lorentz-Kraft Elektronen und Löcher und treibt sie auf die jeweiligen Seitenflächen des Quaders. Bei Kurzschluß fließt ein Strom, dem Informationen über Rekombinationsmechanismen entnommen werden können. Die beschriebene Erscheinung wird als *fotoelektromagnetischer Effekt (PEM-Effekt)* bezeichnet.

8.1.6.4. Lumineszenzdioden. Unter Lumineszenz wird die Lichtemission nach vorangegangener Anregung durch Energieabsorption verstanden. Lumineszenz tritt bei Temperaturen weit unterhalb nennenswerter thermischer Strahlungsemission auf und verschwindet bei hinreichend hohen Temperaturen z. B. infolge strahlungsloser thermischer Rekombination über Zwischenzustände. Nach den Anregungsarten der Lumineszenz in Festkörpern unterscheidet man Foto-, Elektro- oder Katodolumineszenz, sowie Tribolumineszenz (bei mechanischer Deformation z. B. an Zinksulfid) und Chemilumineszenz (im Verlaufe chemischer Reaktionen auftretend, z. B. als Leuchten weißen Phosphors im Dunkeln infolge Oxidation).

Bei Fotolumineszenz ist die Energie der absorbierten Photonen i. allg. größer als die Energie der emittierten Photonen. Ausnahmen sind Zwei-Photonenanregung mit kohärentem Laserlicht oder Mehrstufenanregung. Katodolumineszenz erfolgt bei Anregung des Festkörpers durch von außen auftretende energiereiche Elektronen. Anwendungsbeispiele sind die Oszillografen- und Fernsehbildröhre. Die für das Entstehen von Elektrolumineszenz nötigen hohen elektrischen Feldstärken können entweder von außen angelegt werden (z. B. beim Destriau-Effekt durch ein äußeres elektrisches Wechselfeld) oder als innere Felder an Kontakten und *pn*-Übergängen vorhanden sein. Für die technische Nutzung dominiert die Anregung durch Injektion von Minoritätsladungsträgern über *pn*-Übergänge. Diese Anregungsart liegt den weit verbreiteten Lumineszenzdioden zugrunde. Die Strahlungsemission entsteht dann als eine Folge spezieller Rekombinationsprozesse, die als Umkehr der Absorptionsprozesse, bei denen die Energie der absorbierten Photonen auf Ladungsträger übertragen wird, aufzufassen sind.

Die mit den Rekombinationsprozessen verbundenen Übergänge lassen sich mit dem Energiebändermodell verstehen. Typische Übergangsformen sind Band-Band-Übergänge, Band-Störstellen (bzw. Störstellen-Band-) Übergänge und Übergänge zwischen Störstellen (z. B. zwischen einem Donator und einem Akzeptor).

Bei Band-Band-Übergängen gleicht die Photonenenergie $h\nu$ ungefähr dem energetischen Abstand W_L des oberen Valenzbandrandes vom unteren Leitungsbandrand. Das gilt jedoch nur, wenn das Elektron beim Übergang seinen Impuls $\boldsymbol{p} = \hbar \boldsymbol{k}$ ($\boldsymbol{k}$ = Wellenzahlvektor) nicht ändert. Übergänge dieser Art heißen direkte Übergänge. Die Konstanz des Elektronenimpulses bedeutet, daß Gitterschwingungen (im Aspekt von Quasiteilchen: Phononen) an dieser Art des Rekombinationsprozesses nicht beteiligt sind. Indirekte Übergänge vollziehen sich in zwei Stufen. Die erste Stufe führt innerhalb des Leitungsbandes zur Emission (oder Absorption) eines Phonons, d. h. zum Impulsaustausch des Elektrons mit dem Kristallgitter. Die zweite Stufe ist ein direkter Übergang des Elektrons vom Leitungsband in das Valenzband zur Rekombination mit einem Defektelektron.

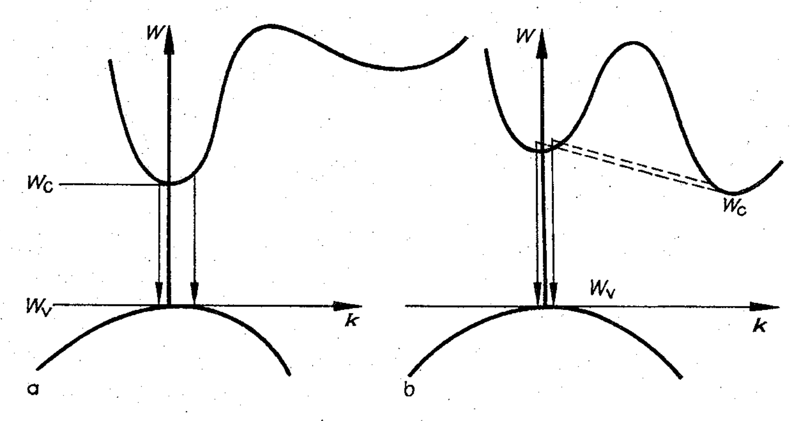


Abb. 8.1.6–5 Band-Band-Rekombination bei (a) direktem und (b) indirektem Übergang

Der Zwischenzustand zwischen beiden Teilprozessen ist virtuell. Der Energieerhaltungssatz muß für den Gesamtprozeß gelten, der Impulserhaltungssatz für jede Stufe einzeln. Wählt man für die Abszisse der Energiebanddarstellung den Wellenzahlvektor k, der dem Elektronenimpuls proportional ist ($p = \hbar k$), lassen sich direkte und indirekte Übergänge zwischen zwei Energiebändern wie in Abb. 8.1.6–5 darstellen. Ob sich in einem Halbleitermaterial bei Absorption oder Emission von Strahlung direkte oder indirekte Übergänge vollziehen, hängt von der Energiebandstruktur des Materials ab. Galliumarsenid (GaAs) gehört zu den direkten, Germanium, Silicium und Galliumphosphid (GaP) gehören zu den indirekten Materialien. Die Wahrscheinlichkeiten für das Stattfinden sind für indirekte Übergänge kleiner als für direkte, so daß im Prinzip direkte Halbleiter besser zur Herstellung von Lumineszenzdioden geeignet sind. Für indirekte Halbleiter läßt sich die Ausbeute erhöhen, wenn bei geeigneter Dotierung die Übergänge von oder in Störstellen erfolgen können.
Für die technisch genutzten Halbleitermaterialien sind strahlende Übergänge unter Beteiligung von Störstellen wichtig.
Eine Auswahl möglicher Fälle zeigt Abb. 8.1.6–6. Die Neutralisierung ionisierter Donatoren und Akzeptoren in *a* liefert wegen der vergleichsweise geringen freiwerdenden Energie Strahlung im fernen Infrarot. Die Frequenz der Strahlung, die von Übergängen zwischen Leitungsband und Akzeptoren bzw. von Donatoren zum Valenzband (Abb. 8.1.6–6b) verursacht wird, ist im Fall direkter Übergänge $\nu = (W_L - W_i)/h$, im Fall indirekter Übergänge $\nu = (W_L - W_i - W_p)/h$ (W_L = Energieabstand des unteren Leitungsbandrandes vom oberen Valenzbandrand, W_i = Energiedifferenz der Störstelle zum nächstgelegenen Bandrand, W_P = Energie des bei der Rekombination emittierten Phonons).
Im Fall c findet die Rekombination zwischen zwei räumlich fixierten Störstellen statt. In die frei werdende Energie ist hier die abstandsabhängige Energie der Coulomb-Wechselwirkung einzubeziehen. Die zu erwartende Strahlungsfrequenz ist $\nu = (W_L - W_A - W_D + e^2/[4\pi\varepsilon r])/h$ (W_A, W_D = Störstellenenergie des Akzeptors bzw. Donators, e = Elementarladung, ε = Dielektrizitätskonstante, r = Abstand Donator – Akzeptor). Da die Störstellen als substituierte Atome auf Gitterplätzen sitzen, sind ihre möglichen Abstände r diskrete Werte (Abb. 8.1.6–7). Bei einer solchen Donator-Akzeptor-Rekombination können im Lumineszenzspektrum bei hinreichend tiefen Temperaturen die Linien, die zu den unterschiedlichen Donator-Akzeptor-Abständen gehören, aufgelöst werden. Dabei sind die Übergangswahrscheinlichkeiten zwischen Paaren geringeren Abstands größer als zwischen weiter entfernteren, d. h. die Strahlungsintensität nimmt mit kleiner werdendem Paarabstand zu.
Übergänge der Art d stellen Beispiele strahlungsloser Rekombination, Auger-Übergänge, dar, die zur Erzielung eines großen Wirkungsgrades (z. B. Verhältnis der optischen Ausgangsleistung zur elektrischen Eingangsleistung) nur von geringem Einfluß sein dürfen. Hier wird die bei der Rekombination freiwerdende Energie einem Elektron als kinetische Energie übertragen.
Eine spezielle Störstellenart entsteht, wenn ein Atom des Wirtsgitters durch ein anderes, das in derselben Gruppe des Periodensystems steht, ersetzt wird. Solche Störstellen, die man als isoelektronisch bezeichnet, halten Ladungsträger infolge geänderter Abschirmung des Atomkerns fest, während die Bindung von Ladungsträgern an Donatoren und Akzeptoren durch die Coulomb-Kräfte zustande kommt. Beispielsweise bilden Stickstoffatome auf Phosphorplätzen in Galliumphosphid (GaP) isoelektronische Störstellen mit Akzeptorcharakter. An isoelektronische Störstellen können sich auch Exzitonen binden. Exzitonen sind, ähnlich dem Wasserstoffatom, durch Coulomb-Kräfte aneinander gebundene Elektron-Loch-Paare, die im freibeweglichen Zustand Energie, aber keine resultierende Ladung transportieren. Wenn an isoelektronische Störstellen gebundene Exzitonen rekombinieren, ergeben sich im Emissionsspektrum schmale Linien.
An Lumineszenzdioden treten meist mehrere Rekombinationsmechanismen gleichzeitig auf. Die emittierte Strahlungsintensität ist der Zahl der einzelnen Rekombinationsakte proportional. Um in der praktischen Realisierung viele Ladungsträgerpaare zur Verfügung zu haben, findet die Rekombination in einem in Flußrichtung vorgespannten *pn*-Übergang statt, in den dadurch zusätzliche Minoritätsträger aus dem jeweils hochdotierten *n*- bzw. *p*-Gebiet injiziert werden (Injektionslumineszenz). Abb. 8.1.6–8a zeigt das hochdotierte *n*- und *p*-Gebiet. Die Gebiete werden als „entartet“ bezeichnet, weil die Fermi-Energie, die Energie bei

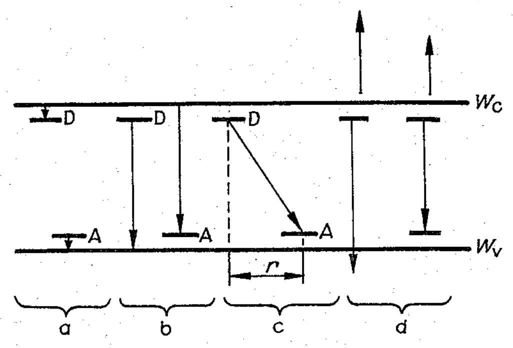


Abb. 8.1.6–6 Über Störstellen ablaufende Rekombinationsmechanismen, (a) Neutralisation ionisierter Donatoren D und Akzeptoren A, (b) „tiefe“ Übergänge Donator-Valenzband bzw. Leitungsband-Akzeptor, (c) Donator-Akzeptor-Übergang, (d) Fälle strahlungsloser (Auger-)Übergänge

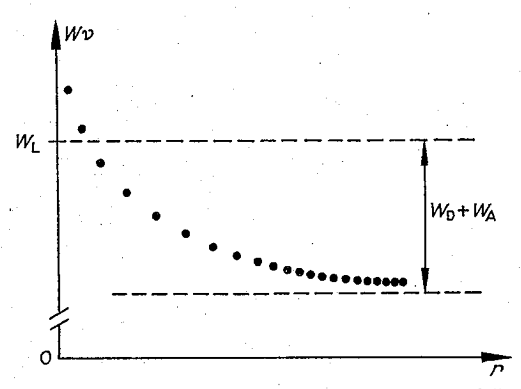


Abb. 8.1.6–7 Einfluß der Coulomb-Wechselwirkung auf die Strahlungsenergie bei Donator-Akzeptor-Paar-Rekombination, r = Donator-Akzeptor-Abstand

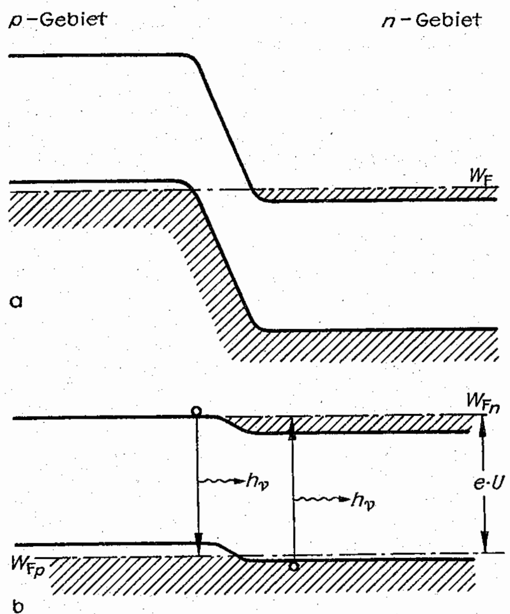


Abb. 8.1.6–8 Ladungsträgerinjektion und -rekombination im *pn*-Übergang, (a) im thermischen Gleichgewicht, (b) bei angelegter Flußspannung U(W_F = Fermi-Energie; W_{F_p}, W_{F_n} = Quasifermi-Energie im *p*- bzw. *n*-Gebiet)

der die Besetzungswahrscheinlichkeit eines Zustandes 1/2 ist, innerhalb des Valenzbandes (im *p*-Gebiet) bzw. innerhalb des Leitungsbands (im *n*-Gebiet) liegt. Das ist für Metalle, nicht aber für Halbleiter typisch. Durch Vorspannung (≈ 2 V) des *pn*-Übergangs in Flußrichtung (Abb. 8.1.6–8b) setzt die Rekombinationsstrahlung ein. Bis auf wenige Ausnahmen bestehen Lumineszenzdioden aus $A_{III}B_V$-Verbindungen. GaAs-Dioden strahlen im Infraroten. Bei Zinkdotierung liegt das Maximum des Emissionsspektrums bei 900 nm (Tab. 8.1.6–9). Bei Siliciumdotierung bilden sich bei Substitution auf Galliumplätzen Donatoren, bei Einbau auf Arsenplätzen Akzeptoren. Das Maximum des Emissionsspektrums von siliciumdiffundierten GaAs-Lumineszenzdioden liegt bei 940 nm. Durch zweckmäßige Wahl der Dotierungselemente erweist sich das indirekte Halbleitermaterial Galliumphosphid (GaP) als vielseitig anwendbar. An Stickstoffatomen als isoelektronische Störstellen werden Exzitonen gebildet, bei deren Zerfall mit 2,22 eV grüne Strahlung auftritt. Bei höherer Stickstoffkonzentration verringern sich die Exzitonenbindungsenergien und das Emissionsmaximum verschiebt sich ins Gelbe. Mit Zink und Sauerstoff dotierte GaP-Dioden emittieren rote Strahlung. Für blau leuchtende Dioden, die im Labor aus Galliumnitrid (GaN) hergestellt werden (W_L = 3,5 eV) gibt es noch keine industriell ausgereifte Herstellungstechnologie. Auch das geeignete Siliciumcarbid (SiC) ist technologisch schwer beherrschbar. Als geeignete Materialien zur Herstellung von Lumineszenzdioden behaupten sich auch ternäre III-V-Verbindungen als Substitutionsmischkristalle des Typs $A^{III_1}_{1-x}\,A^{III_2}_{x}\,B^{V}$ oder $A^{III}\,B^{V_1}_{1-x}\,B^{V_2}_{x}$. Sie sind ebenfalls für die Herstellung von Halbleiter-Laserdioden von Interesse. Da bei diesen Substanzen der Bandabstand eine Funktion des Mischungsparameters x ist, kann mit ihm das Emissionsmaximum verändert werden. Für $GaAs_{1-x}P_x$-Dioden verschiebt sich der Bandabstand W_L von Galliumarsenid (GaAs) mit 1,43 eV (direkter Übergang) zu Galliumphosphid (GaP) auf 2,26 eV (indirekt). Genutzt werden weiterhin die ternären Substitutionsmischkristallreihen $Ga_{1-x}Al_xAs$, $I_{n1-x}Al_xP$ und $In_{1-x}Ga_xP$.

Entsprechend dem Anwendungszweck haben sich im wesentlichen drei Bauarten durchgesetzt, Flächen-, Kugel- und Kantenstrahler (Abb. 8.1.6–10). Flächenstrahler geben das Licht senkrecht zur Fläche des nur wenig unter der Oberfläche liegenden strahlenden *pn*-Übergangs ab. Die Fläche wird meist mit einer Vergußlinse aus Acrylpolyester versehen. Das vergrößert optisch die emittierende Fläche und erhöht die axiale Strahlungsdichte. Diese Bauform eignet sich auch bei GaAs-Lumineszenzdioden, die zur optischen Signalübertragung bis in den MHz-Bereich dienen sollen. An sie kann durch ein geeignetes Vergußmittel ein Lichtleiter angekoppelt werden (Burrus-Diode). Aufwendiger ist es, über dem strahlenden *pn*-Übergang eine halbkugelförmige Domstruktur anzubringen, die eine günstigere Strahlungsverteilung bewirkt, wenn für geringe Eigenabsorption gesorgt ist. Mit dem Kantenstrahler wird auf begrenzter Fläche eine hohe Leuchtdichte erzielt. Der Kantenstrahler ist in seinem Aufbau mit dem Halbleiterlaser identisch. Typische Maße des Diodenmaterials sind 0,5 mm × 0,5 mm × 0,15 mm. Neben der Anwendung zur optischen Signalübertragung im Infraroten über Lichtleiter werden Lumineszenzdioden auch für Lichtschranken, im sichtbaren Bereich zur Anzeige von Schalt- und Betriebszuständen und für Prüfeinrichtungen eingesetzt. Wegen ihrer kurzen Ansprechzeit können sie auch zur Anzeige von alphanumerischen Zeichen benutzt werden. Bei kleinen Dimensionen ist eine monolithische Fertigung möglich (Matrix mit sieben Zeilen zu je fünf Lichtpunkten, oder zur Zifferndarstellung als 7-Segment-Anordnung). Alterungserscheinungen treten nur bei elektrischer und thermischer Überlastung auf. Lumineszenzdioden haben gegenüber Glühlampen eine außerordentlich große Zuverlässigkeit. Typische Lebensdauern liegen bei etwa 10^5 Stunden (≈ 10 Jahre).

Tab. 8.1.6–9 Lumineszenzdioden, Substanzen und Wellenlängen des Emissionsmaximums

Substanz	Wellenlänge in nm	Farbe
GaN	450	blau
GaP : N	565	grün
$GaAs_{0,15}P_{0,85}$: N	590	gelb
$GaAs_{0,3}P_{0,7}$: N	630	orange
$GaAs_{0,6}P_{0,4}$	650	rot
GaP : Zn, O	690	rot
GaAs : Zn	900	infrarot

8.1.6.5. Laserdioden. Mit Anordnungen, die Lumineszenzdioden (Abb. 8.1.6–10c) analog sind, kann bei höheren Flußstromstärken, von einer bestimmten Schwellstromdichte an, Lasertätigkeit einsetzen. Die Art ihrer Herbeiführung unterscheidet den Halbleiterlaser von den anderen Festkörperlasern (z. B. Rubinlaser mit dem Wirkmaterial Aluminiumoxid (Al_2O_3) und Cr^{3+} als aktiven Ionen) und von Flüssigkeits- und Gaslasern. Die strahlenden Übergänge erfolgen im Halbleiter nicht zwischen diskreten Energiezuständen von Elektronen, sondern zwischen Energiebändern, dem Leitungsband und dem Valenzband. Die

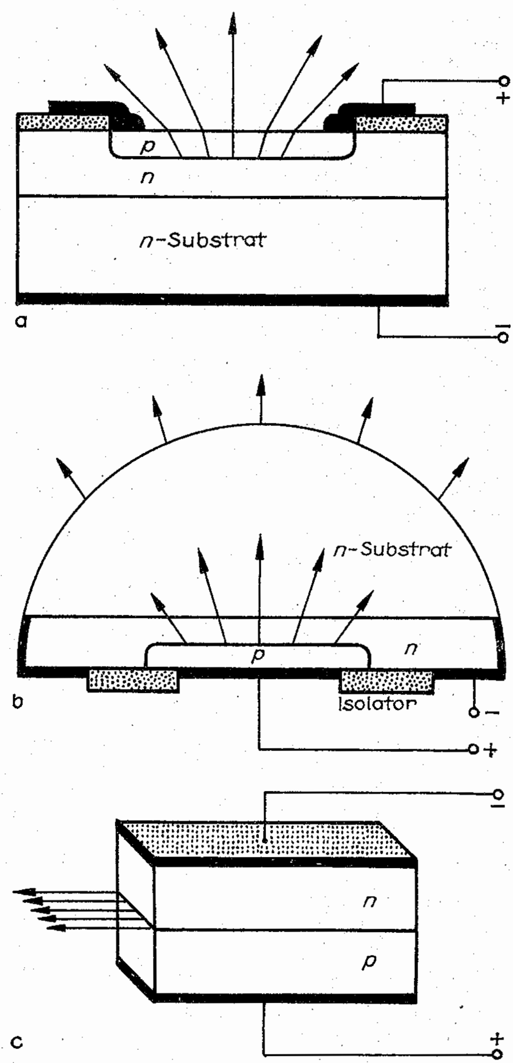
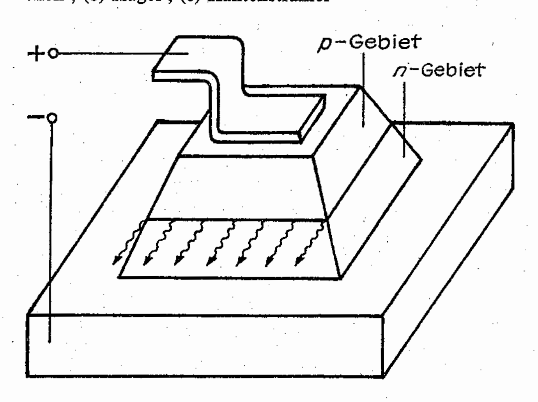


Abb. 8.1.6–10 Bauformen von Lumineszenzdioden, (a) Flächen-, (b) Kugel-, (c) Kantenstrahler



spontane Strahlung, d.h. die allgemein beobachtete Rekombinationsstrahlung in Lumineszenzdioden, erfüllt deshalb auch noch ein relativ breites Frequenzintervall (bei Galliumarsenid (GaAs) z.B. 30 nm). Ein Laser (engl. *L*ight *a*mplification by *s*timulated *e*mission of *r*adiation, Lichtverstärkung durch induzierte Strahlungsemission) ist im Grunde ein Oszillator für optische Frequenzen. Für sein Funktionieren ist die Herbeiführung einer Besetzungsinversion Voraussetzung, d. h. einer Situation, bei der Elektronenzustände höherer Energie zu einem größeren Anteil besetzt sind, als Zustände mit geringerer Energie. Ein solcher Zustand stellt sich nicht automatisch ein, da er weitab vom Gleichgewicht liegt. Die Herbeiführung dieses Zustandes wird bei fast allen anderen Lasergruppen vorwiegend durch optisches Pumpen erreicht. Dem System wird Energie in Form von Strahlung zugeführt, die die Besetzungsinversion hervorruft. Im Halbleiterlaser wird die Besetzungsinversion, d. h. der Elektronenüberschuß im Leitungsband, durch das von außen angelegte elektrische Feld in Flußrichtung erreicht (vgl. Abb. 8.1.6–8). Im *p*-Gebiet entsteht im Leitungsband ein großer Elektronenüberschuß und im *n*-Gebiet im Valenzband ein ebensolcher Defektelektronenüberschuß. Diese Methode der Erzielung der Nichtgleichgewichtsverteilung (der man in der theoretischen Behandlung formal eine negative Temperatur zuordnen kann) bezeichnet man als Ladungsträgerinjektion am *pn*-Übergang. Der Halbleiterlaser heißt deshalb auch Injektionslaser. Nachdem die Elektronen ins Leitungsband gekommen sind, sind folgende Prozesse möglich:

Die quasifreien Elektronen verringern ihre Energie und nähern sich dem unteren Rand des Leitungsbandes innerhalb einer typischen Zeit t_1. Zum anderen können Elektronen spontan, d. h. ohne äußere Einwirkungen, vom Leitungs- in das Valenzband übergehen und mit Defektelektronen rekombinieren und zwar in einer charakteristischen Zeit t_2. Zur Aufrechterhaltung der Besetzungsinversion am unteren Rand des Leitungsbandes ist die Bedingung $t_1 \ll t_2$ zu erfüllen.

Damit ist dann die Voraussetzung gegeben, daß die für die Lasertätigkeit entscheidende stimulierte Emission möglich ist. Es tritt eine Lichtverstärkung durch die im Strahlungsfeld des Lasers vorhandenen Photonen der Laserfrequenz ein. Aufgrund der Besetzungsinversion überwiegt die Intensitätsvergrößerung infolge stimulierter Emission, die Intensitätsverringerung infolge der Absorption von Photonen auf Kosten einer Elektronenanregung. Das kommt einem negativen Wert des Absorptionskoeffizienten und damit einem Anwachsen der Lichtintensität längs dem *pn*-Übergang gleich.

Im Halbleiterlaser läuft das Licht zwischen zwei parallelen Spiegelflächen, die senkrecht zum *pn*-Übergang liegen und als Resonatoren dienen, hin und her (Abb. 8.1.6–11) analog der vom Interferometer nach Perot und Fabry bekannten Anordnung. Die im Resonatorraum ausgelösten Lichtquanten führen den Verstärkungseffekt über Rückkopplung dadurch herbei, daß sie neue Rekombinationsakte stimulieren und über die Resonatorlänge L für die Entstehung einer kohärenten Strahlung sorgen. Durch die vorgegebene Länge des Resonatorraums können nur solche Schwingungen verstärkt werden, deren halbe Wellenlänge ein ganzzahliges Vielfaches der Resonatorlänge ist. Der Spektralbereich, in

Abb. 8.1.6–11 Schema einer Laserdiode, Vorder- und Rückfläche sind zueinander parallel und erfüllen die Funktion der Resonatorspiegel

dem eine Lichtverstärkung möglich ist, hat eine Breite von ungefähr kT (k = Boltzmann-Konstante, T = absolute Temperatur).
Bei einer Resonatorlänge L von einigen 100 bis 1000 Wellenlängen kann es daher zur Ausbildung mehrerer Schwingungsmoden kommen, $m\lambda/2 = nL$ ($m = 1, 2, \ldots$) (n = Brechungsindex des Halbleiters, λ = Vakuumwellenlänge).
Das Licht verläßt das Bauelement durch die partiell durchlässige Resonatorfläche. Die senkrecht dazu liegenden Flächenpaare werden optisch rauh gehalten, so daß Lichtverstärkung nicht wirksam wird und in Richtungen, die aus der Ebene des *pn*-Übergangs herausführen, überwiegt die Absorption und verhindert eine Lasertätigkeit.
Die Schwellstromdichte, bei der der Übergang zur induzierten Emission erfolgt, liegt bei tiefen Temperaturen für eine GaAs-Laserdiode mit 0,4 mm Resonatorlänge und einer 1 µm dicken aktiven Zone am *pn*-Übergang bei 400 A/cm². Bei höheren Temperaturen steigt die Schwellstromdichte wegen zunehmender Absorptionsverluste durch freie Ladungsträger an. Nach der Theorie des Halbleiterlasers lassen sich die Schwellstromdichten durch die Vergrößerung der Resonatorlänge, die Verringerung der Dicke der aktiven Zone sowie durch die reflektierende Beschichtung der Spiegelflächen herabsetzen. Infolge der hohen thermischen Belastung gibt es gegenwärtig nur wenig Typen von Halbleiterlasern für Dauerstrichbetrieb bei Raumtemperatur. Die meisten Halbleiterlaser arbeiten im Impulsbetrieb und/oder werden gekühlt. Vorteilhaft ist ihr hoher Wirkungsgrad, der an 100 % heranreichen kann und die einfache Modulationsmöglichkeit der emittierten Strahlung über den Injektionsstrom. Die Laserdiode genügt für die optische Nachrichtenübertragung über Lichtleiter infolge der gegenüber Lumineszenzdioden um mehr als eine Größenordnung verbesserten Monochromasie ($\Delta\lambda \approx 2$ nm gegenüber $\Delta\lambda \approx 40$ nm bei der Lumineszenzdiode [LED]) erhöhten Anforderungen. Darüber hinaus zeigt die Laserdiode eine kürzere Impulsanstiegszeit (0,3 bis 1 ns gegenüber 3 bis 10 ns bei der LED). Wegen der besseren Strahlungsbündelung kann von einem Diodenlaser ein höherer Anteil der Lichtleistung in einen Lichtleiter eingekoppelt werden.
Noch nicht voll beherrscht werden die Alterungserscheinungen von Laserdioden. Für die allmähliche Verringerung der emittierten Strahlungsintensität scheint die Wirkung von im Verlauf des Betriebs in der aktiven Schicht entstehenden Kristallbaufehlern verantwortlich zu sein.
Als Materialien für Laserdioden haben sich vorwiegend die auch für Lumineszenzdioden bewährten $A_{III}B_V$-Verbindungen erwiesen, wobei das technologische Interesse der Verringerung der Breite der aktiven Zone und gleichzeitig der optimalen Anpassung der Gitterkonstanten der am *pn*-Übergang aufeinandertreffenden Materialien ist. Bevorzugt wird das System $Ga_{1-x}Al_xAs$, dessen binäre Endglieder, das Aluminiumarsenid (AlAs) und das Galliumarsenid (GaAs) nahezu die gleiche kristallografische Gitterkonstante aufweisen.

8.1.6.6. Fototransistoren haben eine gegenüber Fotodioden um den Faktor der Stromverstärkung des Transistors erhöhte Empfindlichkeit bei Beleuchtung. Der Basis-Kollektor Übergang des *Bipolar-Fototransistors* ist dem einfallenden Licht zugänglich. Fototransistoren haben meist keine herausgeführte Basiselektrode. Wie bei der Fotodiode werden die durch den inneren Fotoeffekt erzeugten Elektron-Defektelektron-Paare durch das Feld des Raumladungsgebietes im Basis-Kollektor-*pn*-Übergang getrennt. Betrachtet man einen *npn*-Fototransistor (Abb. 8.1.6–12), so entsteht ein von den Minoritätsladungsträgern der Basis, den Elektronen, getragener Sperrstrom in den Kollektor. Dieser belichtungsabhängige Strom muß eine Änderung des Emitterstromes verursachen, da sich der Basisstrom als Differenz zwischen Emitter- und Kollektorstrom darstellt. Entsprechend der Stromverstärkung des Transistors in Emitterschaltung besteht zwischen dem durch Lichteinfall erzeug-

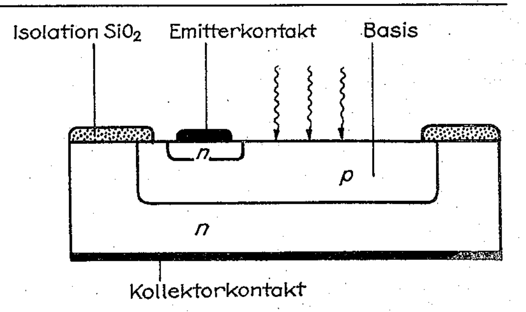


Abb. 8.1.6–12 *npn*-Fototransistor (schematisch)

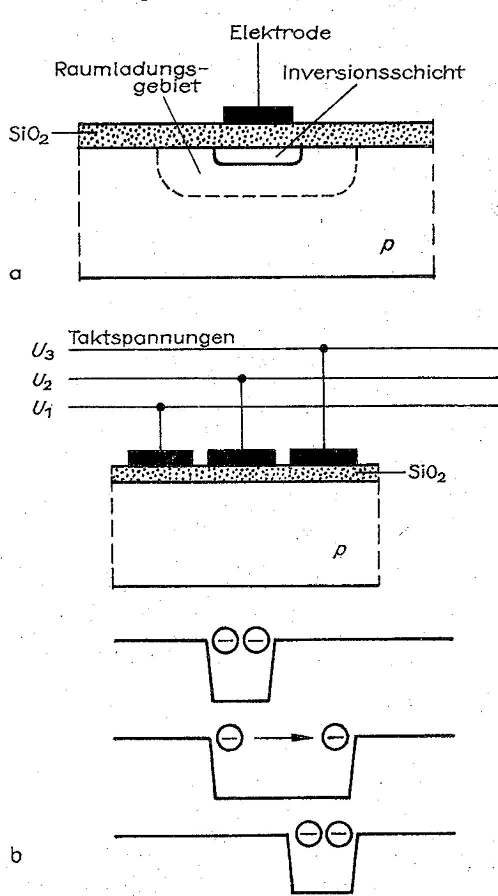


Abb. 8.1.6–13 Ladungsgekoppeltes Bauelement (CCD), (a) MOS-Kondensator, (b) Betrieb eines dreiphasigen ladungsgekoppelten Bauelements (schematisch)

ten Fotostrom in der Basis I_f und dem Kollektorstrom I_c die Beziehung $I_c=(h_{21}^{(e)}+1)\ I_f$ ($h_{21}^{(e)}$=Stromverstärkung bei Emitterschaltung).
Fokussiert man bei lichtempfindlichen Sperrschichtfeldeffekttransistoren die Strahlung auf das Gebiet der Gate-Elektrode, so ändert sich der Drain-Strom. Die *Foto-Feldeffekttransistoren* (Foto-FET) haben sehr kleine Dunkelströme und sind bei extrem flach liegendem steuernden *pn*-Übergang als empfindliche Halbleiterdetektoren im ultravioletten Spektralbereich einsetzbar. Eine Abwandlung des Feldeffekttransistors ist ein MOS-(Metall-Oxid-Halbleiter)-Kondensatorelement (Abb. 8.1.6–13). Auf einem *p*-leitenden Substrat befindet sich eine Isolationsschicht aus Siliciumdioxid (SiO_2) mit einer Metallelektrode, die auf positivem Potential liegt. Wenn die Unterlage an Masse liegt, sammeln sich an der Metallelektrode Elektronen, die die Akzeptoratome negativ aufladen und eine Raumladung bilden, die speicherbar ist.
Werden diese Kondensatorelemente zu einer Zeilen- oder Matrixstruktur mit geringen Zwischenabständen von wenigen Mikrometern aneinandergefügt, so daß ihre Raumladungen ineinander übergreifen, dann können mit Hilfe einer mehrphasigen Steuerspannung die Ladungen von Kondensator zu Kondensator einer Reihe verschoben werden (Abb. 8.1.6–13) und stehen nacheinander, nachdem sie aufgenommen wurden, zur Weiterverarbeitung zur Verfügung. Anordnungen dieser Art werden als *ladungsgekoppelte Bauelemente* (engl. *c*harge *c*oupled *d*evice [CCD]) bezeichnet.
Die Aufladung dieser MOS-Kondensatoren erfolgt entweder durch die Eingabe elektrischer Signale am Eingang einer Zeile, wodurch Speicher in der Art von Schieberegistern entstehen oder durch Lichteinwirkung. Bestrahlt man ein solches Kondensatorelement, so ist die gespeicherte Ladung der Zahl der absorbierten Photonen proportional. Wenn durch geeignete Maßnahmen verhindert wird, daß während des Ladungstransports durch Belichtung neugebildete Ladungen die in den transportierten enthaltene Informationen verwischen, ist eine Matrix von ladungsgekoppelten Bauelementen als *Halbleiter-Bildaufnahmeeinheit* verwendbar. Es wird erwartet, daß Halbleiter-Bildaufnahmeeinrichtungen, die sich noch im Versuchsstadium befinden, die Bildaufnahmeröhren mit Elektronenstrahlabtastung verdrängen werden. Als Sensoren haben sie bei der Automatisierung von Produktionsprozessen vielfältige Anwendungsmöglichkeiten bei der Erkennung von Gegenständen, Bildern und Zeichen, bei der Messung von Winkeln, Längen und Volumina sowie bei der Ermittlung von energieabstrahlenden Objekten.
Auch Thyristoren lassen sich auf optischem Wege beeinflussen. Der bei einem gesteuerten Gleichrichter normalerweise durch einen Strom über den Torkontakt herbeigeführte Übergang vom Blockier- in den Durchlaßzustand läßt sich durch Elektron-Defektelektron-Paarerzeugung mit Hilfe einfallenden Lichts vollziehen. Dazu wird der mittlere der drei *pn*-Übergänge dem Licht zugänglich gemacht. Die Ansteuerung dieser aus Silicium hergestellten *Fotothyristoren* erfolgt mit Lumineszenzdioden oder Halbleiterlasern.

8.1.6.7. Optoelektronische Koppler. In optoelektronischen Kopplern unterliegt das am Eingang liegende elektrische Signal einer Umwandlung in eine optische Form. Am Ausgang erscheint es wieder in elektrischer Form. Vorteile solcher Anordnungen sind eine vollständige galvanische Trennung von Eingang und Ausgang und damit Rückwirkungsfreiheit. Optoelektronische Koppler bestehen aus einem Strahlungssender, meist einer im Infraroten emittierenden GaAs-Lumineszenzdiode, einem optischen Übertragungsmedium, z. B. einem Lichtleiter und einem Lichtempfänger. Der Lichtempfänger ist entweder eine Fotodiode (Abb. 8.1.6–14), eine Schottky-Fotodiode oder ein Fototransistor. Auch andere Empfänger sind möglich. Durch die Wahl von Silicium als Grundmaterial des Empfängers wird eine günstige Übereinstimmung der spektralen Verteilung der Strahlung des Senders und der spektralen Empfindlichkeit des Empfängers gewährleistet. Die konstruktive Ausführung der Koppelelemente kann auch eine äußere Beeinflussung des Strahlengangs zwischen Sender und Empfänger vorsehen (offene Koppelelemente). So enthält z. B. der optoelektronische Gabelkoppler zwischen Sender und Empfänger einen Luftspalt zur Einführung eines Unterbrechers. Optoelektronische Koppler werden sowohl analog als auch digital betrieben. Sehr kurze Ansprechzeiten zwischen 10^{-9} und 10^{-10} s sind erreichbar, wenn als Strahlungsempfänger Lawinenfotodioden oder Schottky-Fotodioden eingesetzt werden. Diese Kombinationen sind dem digitalen Betrieb vorbehalten.

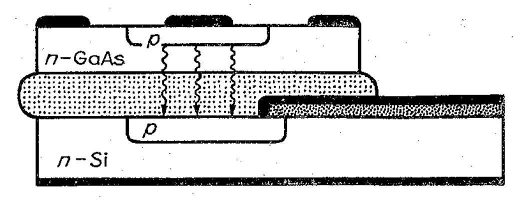


Abb. 8.1.6–14 Optoelektronischer Koppler mit GaAs-Lumineszenzdiode als Lichtsender und einer Silicium-Fotodiode als Lichtempfänger (schematisch)

8.1.6.8. Lichtleiter. Das ständige Anwachsen von zu übertragenden Informationsmengen hat einen steigenden Bedarf an Nachrichtenkanälen zur Folge, verbunden mit dem Bestreben, den Informationsfluß, d. h. die je Zeiteinheit übertragene Informationsmenge eines Kanals zu vergrößern. Diesen Forderungen wird das Licht als Informationsträger in idealer Weise gerecht. Die Nutzung der Vorzüge des Lichtes ist jedoch an drei Voraussetzungen gebunden: 1. müssen die Informationen in Lichtsignale umgewandelt werden, 2. muß aus den empfangenen Signalen die Information in ihre Ausgangsform zurückgeführt werden und 3. darf der Übertragungskanal Störquellen nur in einem solchen Umfang zulassen, daß eine verfälschte oder unvollständige Übertragung nahezu ausgeschlossen ist. Als einen ersten Versuch der optischen Nachrichtenübertragung kann man Lichtsprechgeräte ansehen, die unter Nutzung des nahen Infrarot bei 0,8 μm mit Hilfe thermischer Lichtquellen Sprachsignale in offener Atmosphäre über wenige Kilometer übertragen konnten.
Mit Beginn der 70er Jahre verbesserten sich zunehmend die Voraussetzungen dafür, die Vorteile des Lichtes zur Informationsübertragung nutzen zu können. Die erste Voraussetzung kann durch die Bereitstellung spezieller Lichtquellen als erfüllt betrachtet werden. Temperaturstrahler, die nahezu über den gesamten sichtbaren Bereich und einen großen Teil des infraroten Anteils des elektromagnetischen Spektrums emittieren und sich nur bis zu einigen Kilohertz modulieren lassen, sind damit für diesen Zweck bedeutungslos geworden. An ihre Stelle traten die Halbleiterbau-

elemente Lumineszenzdiode (lichtemittierende Diode, LED) und Laserdiode sowie andere Lasertypen. Lumineszenzdioden (vgl. 8.1.6.4.), die bei typischen Strömen um 10 mA Lichtströme um 10 mlm aussenden, emittieren zwar inkohärentes Licht, aber das mit relativ guter Monochromasie ($\Delta\lambda \approx 30$ bis 50 nm). Der Wirkungsgrad der Energieformwandlung, das Verhältnis der optischen Ausgangsleistung zur elektrischen Eingangsleistung erreicht Maximalwerte um 10 %. Da die Strahlungsleistung dem Diodenstrom proportional ist, ergibt sich daraus bereits eine Modulationsart, die bis in den 100-MHz-Bereich anwendbar ist. Für höhere Forderungen stehen Laserlichtquellen, die monochromatische ($\Delta\lambda \approx 2$ nm) und kohärente Strahlung emittieren, zur Verfügung, unter denen die Halbleiterlaserdiode für die Anwendung von besonderem Interesse ist. Die zweite Voraussetzung, der Empfang der optischen Signale und ihre Rückverwandlung in die Ausgangsinformation, wird durch spezielle Formen von Fotodioden gewährleistet.

Diese Dioden arbeiten im Sperrzustand und der über sie fließende Strom setzt sich aus dem unerwünschten und deshalb klein zu haltenden Sperrstrom und dem Fotostrom zusammen. Der Fotostrom entsteht durch den inneren lichtelektrischen Effekt, bei dem durch Einstrahlung von Photonen im Raumladungsgebiet Ladungsträgerpaare erzeugt werden, die sich durch die von außen angelegte Spannung trennen und damit den Fotostrom darstellen. Besonders gut eignen sich die Lawinen-Fotodioden (Avalanche-Fotodioden). In ihnen erhalten die sich trennenden Ladungsträgerpaare aufgrund der hohen elektrischen Feldstärke im Raumladungsgebiet eine so große Energie, daß sie durch Stoßionisation lawinenartig weitere Ladungsträger freisetzen und damit einen inneren Verstärkungseffekt nutzen. Eine andere Fotodiodenart, die *pin*-Fotodioden (vgl. 8.1.6.2.), enthält zwischen *p*- und *n*-Gebiet eine eigenleitende (*i*ntrinsic), hochohmige Schicht, um die Quantenausbeute zu vergrößern und den Sperrstrom zu verringern. Für die optische Informationsübertragung ist schließlich die dritte Voraussetzung mit den Eigenschaften des Mediums, in dem die Übertragung stattfindet, verknüpft. Dabei ist zu beachten, daß die Eigenschaften des Übertragungsmediums ihrerseits spezielle Forderungen an Sender und Empfänger stellen.

Die naheliegende Möglichkeit, die Atmosphäre als Übertragungsmedium zu nutzen, bringt jedoch so viele Nachteile und schwer zu kalkulierende Schwankungen, daß sie, außer für Sonderfälle geringer Entfernung, nicht genutzt werden kann (Abb. 8.1.6–15). Hinderlich sind die Brechung an Luftschichten unterschiedlicher Dichte, die Schwächung durch Absorption und die Streuung an Schwebeteilchen, sowie die durch thermische Instabilitäten entstehenden atmosphärischen Turbulenzen. Es ist versucht worden, die mit dem Übertragungsmedium Atmosphäre verbundenen Komplikationen zu umgehen. Dazu sollten evakuierte oder mit einem Schutzgas gefüllte Rohrleitungen dienen. Glaslinsen bzw. elektrisch beheizte Rohrstücke („Gaslinsen") sollten die Strahldivergenz korrigieren.

Als technisch nutzbares Übertragungsmedium erwies sich schließlich auch unter aussichtsreichem ökonomischem Aspekt etwa seit 1970 endgültig die Glas- bzw. Kunststoffaser. Faserbündel waren bereits aus der Medizin bekannt, wo sie zur Gewinnung von Abbildern aus dem Körperinnern benutzt werden. In diesen geordneten Faserbündeln hat am Bündelanfang jede Faser die gleichen Nachbarfasern wie am Bündelende. Das sich längs einer Faser ausbreitende Licht wird infolge der Totalreflexion an der Innenwand der Faser immer in der Faser gehalten und verläßt die Faser auch bei Krümmung nicht. Für die Verwendung von Glasfasern als Lichtleiter zum Zwecke der Informationsübertragung sind durch verfeinerte Technologien der Glasherstellung Gläser bereitgestellt worden, die den Anforderungen an eine geringe Schwächung des Lichts durch Absorption und Streuung gerecht werden. Als Schwächungsmaß für die Lichtleistung wird das aus der Nachrichtentechnik bekannte und in Dezibel (dB) anzugebende Dämpfungsmaß benutzt. Es ist der mit zehn multiplizierte dekadische Logarithmus des Leistungsverhältnisses a [dB] $= 10 \log P_1/P_2$. Abb. 8.1.6–16 zeigt die Wellenlängenabhängigkeit der Dämpfung von Lichtleitern zwischen 0,6 µm und 1,8 µm. In der Praxis gelingt es, die Dämpfung zwischen etwa 1,0 µm und 1,8 µm in dem erforderlichen Maße klein zu halten. In Richtung kleinerer Wellenlängen wächst die Dämpfung infolge der Rayleigh-Streuung an Teilchen, die klein gegenüber der Lichtwellenlänge sind. Nach größeren Wellenlängen hin steigt die Dämpfung aufgrund wachsender Absorption der Strahlung und einer damit verbundenen Umsetzung in Wärme. Im Wellenlängengebiet, in dem der Einfluß von Absorption und Streuung etwa gleich groß ist, tauchen noch Dämpfungsspitzenwerte auf, die mit der Konzentration der dämpfungswirksamen OH-Ionen zusammenhängen. Damit sind gegenwärtig für die Anwendung neben dem bereits genutzten Bereich um 0,85 µm zwei Wellenlängenintervalle („Fenster"), $1\,\mu m \leq \lambda \leq 1{,}3\,\mu m$ und $1{,}5\,\mu m \leq \lambda \leq 1{,}8\,\mu m$ verfügbar.

Für den konstruktiven Aufbau von Lichtleitfasern ist wegen der bereit stehenden Lichtquellen von Bedeutung, daß auch noch Licht in die Faser eingekoppelt werden kann, das schon einen relativ großen Winkel mit der Faserachse bildet, aber durch Totalreflexion trotzdem ständig innerhalb der Faser gehalten werden kann. Das ermöglicht weitgehend der Stufenprofillichtleiter (Abb. 8.1.6–17), bei dem sich Kern- und Mantelgebiet durch einen Sprung im Brechungsindex voneinander unterscheiden. Der Sinus des halben Öffnungswinkels $2\,\varphi$, unter dem noch Licht eingekoppelt werden kann, heißt *numerische Apertur* $A_n = \sin\varphi = n_K \sin\varphi = \sqrt{n_K^2 - n_M^2}$.

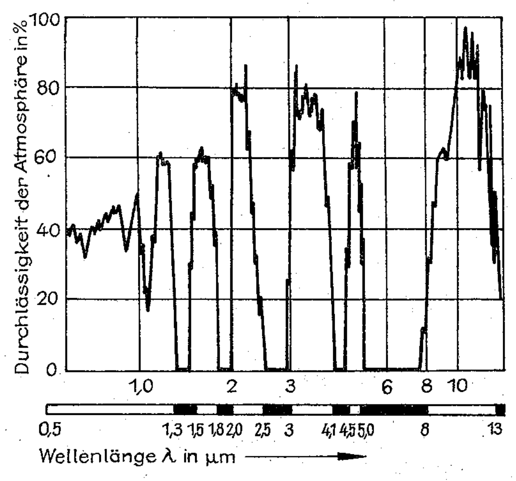


Abb. 8.1.6–15 Transparenz der Atmosphäre für elektromagnetische Wellen zwischen 0,5 µm und 15 µm Wellenlänge

Wenn der Brechungsindex des Kerns $n_K = 1{,}50$ ist, und der des Mantels auf $n_M = 1{,}49$ verringert wird, ergibt das einen Winkel von $\approx 10°$, um den das eingestrahlte Licht von der Faserachse abweichen darf.
Lichtstrahlen, die unter verschiedenen Winkeln in den Lichtleiter eingekoppelt werden, legen jedoch unterschiedlich lange Wege zurück. Achsenparallele Strahlen haben die kürzesten Wege und damit auch die kürzesten Laufzeiten. Strahlen, die unter einem endlichen Winkel zur Faserachse eingestrahlt werden, sind nach mehr oder weniger kurzem Laufweg an der Wand angelangt und werden reflektiert. Schief zur Achse einfallende Strahlen „schrauben" sich durch den Lichtleiter. Die Wege nicht achsenparalleler Strahlen sind länger. Ein in den Lichtleitern eingekoppelter

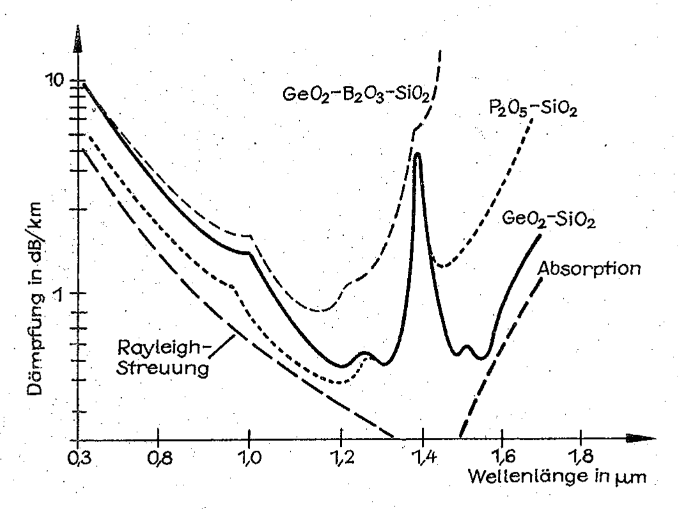


Abb. 8.1.6-16 Dämpfung von Lichtleitfasern in Abhängigkeit von der Wellenlänge

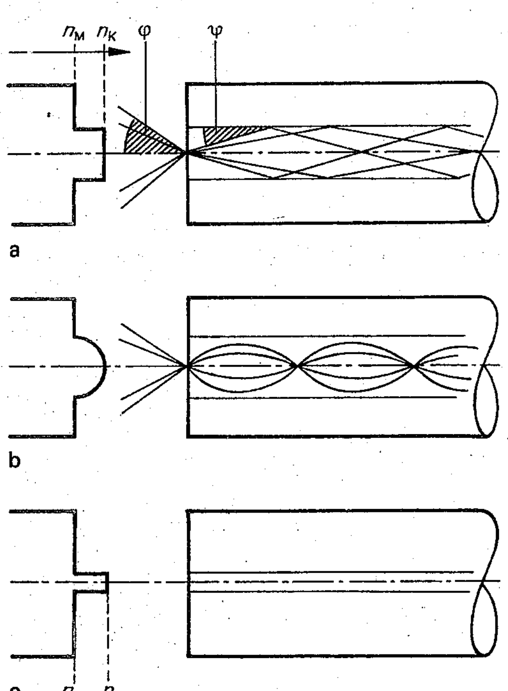


Abb. 8.1.6-17 Verlauf der Brechungsindizes n_K, n_M von Kern und Mantel und der Lichtwege in (a) Stufenprofillichtleiter, (b) Gradientenlichtleiter und (c) Monomodefaser

Lichtimpuls mit endlicher Divergenz wird demnach bei der Auskopplung verbreitert erscheinen. Der zeitlich verbreiterte Ausgangsimpuls ist dabei aus einer Vielzahl von Schwingungsmoden zusammengesetzt, die zeitlich nacheinander am Lichtleiterende ankommen. Diese Erscheinung wird als Modendispersion bezeichnet. Die Modenzahl wächst mit wachsendem Faserdurchmesser an. Als eine Abhilfe gegen die Verbreiterung des Ausgangsimpulses kann die Manipulation des Brechungsindexes als Funktion von Faserradius angesehen werden. Da nicht achsenparallel eingekoppelte Lichtstrahlen infolge ihres geometrisch längeren Wegs im homogenen Lichtleiter auch eine längere Zeit unterwegs sind, muß dafür gesorgt werden, daß sich diese Strahlen mit größerer Geschwindigkeit als die in der Längsachse der Faser ausbreitenden Strahlen bewegen können. Wegen $v = c/n$ (v = Ausbreitungsgeschwindigkeit des Lichts in der Faser, c = Vakuumlichtgeschwindigkeit, n = Brechungsindex) muß der Brechungsindex zum Faserrand hin abnehmen. Diese Eigenschaft hat der Gradientenlichtleiter. Wird die Abhängigkeit des Brechungsindex vom Faserradius optimal eingestellt, kann die Impulsverbreiterung wesentlich reduziert werden. Bei einer Lichtleiterlänge von 1 km kann in Gradientenlichtleitern die Verbreiterung eines schmalen Lichtimpulses kleiner als 1 ns gehalten werden, was Informationsflüsse von mehr als 1 Gbit/s gestattet. Stufenprofillichtleiter und Gradientenlichtleiter sind Multimodenlichtleiter. Ihr Kerngebiet hat einen Durchmesser von ≈ 50 µm und die gesamte Faser ≈ 150 µm.
Reduziert man den Durchmesser des Kerns auf wenige Lichtwellenlängen, so reduziert sich die Modenzahl auf eins, was im strahlenoptischen Modell der ausschließlichen Möglichkeit achsenparalleler Strahlen entspricht und eine weitere Vergrößerung des übertragbaren Informationsflusses bringt. Diese Lichtleiter werden als Monomodelichtleiter (vgl. Abb. 8.1.6-17) bezeichnet. Ihre Kerndurchmesser

liegen zwischen 3 µm...10 µm, wodurch an die Koppeltechnik erhöhte Anforderungen gestellt werden.

Zum zeitlichen Auseinanderlaufen von Lichtimpulsen infolge unterschiedlicher Laufwege im Lichtleiter, der Modendispersion, kommt ein weiterer Grund der Impulsverbreiterung in Form der Abhängigkeit des Brechungsindex des Faserkerns n_K von der Wellenlänge λ. Diese als Materialdispersion bezeichnete Ursache für Laufzeitunterschiede wird durch die spektrale Breite der Lichtquelle beeinflußt.

Zur optischen Verbindung zweier Lichtleiter, die Aufteilung bzw. Verzweigung von Lichtleitern stehen zahlreiche technische Möglichkeiten zur Verfügung, die jedoch weit komplizierter zu handhaben sind, als die Verbindung von Kupferkabeln.

Zur mechanischen Stabilisierung werden Lichtleiterkabel aus mehreren Lichtleitern mit einer gemeinsamen Schutzhülle benutzt.

Für praktisch arbeitende, digitale Lichtleitersysteme müssen die infolge der Dämpfung schwächer gewordenen Signale nach bestimmten Entfernungen durch Verstärker aufgefrischt werden. Die erforderlichen Verstärkerabstände betragen bei Dämpfungen um 5 dB/km ≈ 10 km.

8.1.6.9. Flüssigkristall-Anzeigeelemente. Substanzen, die Flüssigkristalle bilden, bestehen aus langgestreckten Molekülen, die in der Regel zwischen zwei Benzenringen eine Mittelgruppe M und an den entgegengesetzten Enden in Parallelstellung (1,4-Stellung) Flügelgruppen F und F' tragen (Abb. 8.1.6–18).

Die in Flüssigkeitskristallen über größere Bereiche vorhandene Fernordnung und der typische Molekülaufbau als Grundmotiv für die kristalline Struktur führen zu anisotropen Eigenschaften, von denen besonders die optische Anisotropie und das Verhalten in elektrischen Feldern für die praktische Anwendung von Bedeutung sind. Unter den möglichen Typen von Flüssigkristallen weist die nematische Phase mehr oder weniger gut zueinander parallel ausgerichtete Moleküle aus und verhält sich wie ein optisch einachsiger Körper. In Flüssigkristall-Anzeigeelementen wird die optische Aktivität, d. h. die Drehung der Polarisationsebene durch die von außen anliegende elektrische Feldstärke gesteuert und damit zur Anzeige verwendet. Im feldfreien Fall (Abb. 8.1.6–19) liegen die Moleküle parallel zu den ≈ 10 bis 15 µm voneinander entfernten Gefäßwänden. Unter Beibehaltung der Parallelität drehen sich jedoch die Moleküle um eine mit der Gefäßwandnormalen zusammenfallende Achse senkrecht zur Moleküllängsrichtung zwischen beiden Gefäßwänden um 90°. Diese Drehung wird im feldfreien Fall erzwungen. Die Gefäßwände haben aus diesem Grund eine Rillenstruktur, an deren Mulden sich die an den Wänden liegenden Moleküle anpassen. Da diese Rillenstruktur an beiden Gefäßwänden um 90° gegeneinander verdreht ist, gehen die Moleküllagen im Volumen der kristallinen Flüssigkeit kontinuierlich von der einen in die andere Orientierung über. Bei anliegendem elektrischen Feld orientieren sich die Moleküle im Volumen des Flüssigkristalls um und stellen sich nahezu senkrecht zu den

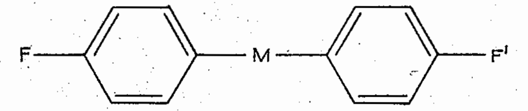


Abb. 8.1.6–18 Allgemeine Konstitutionsformel von Molekülen, die kristalline Flüssigkeiten bilden

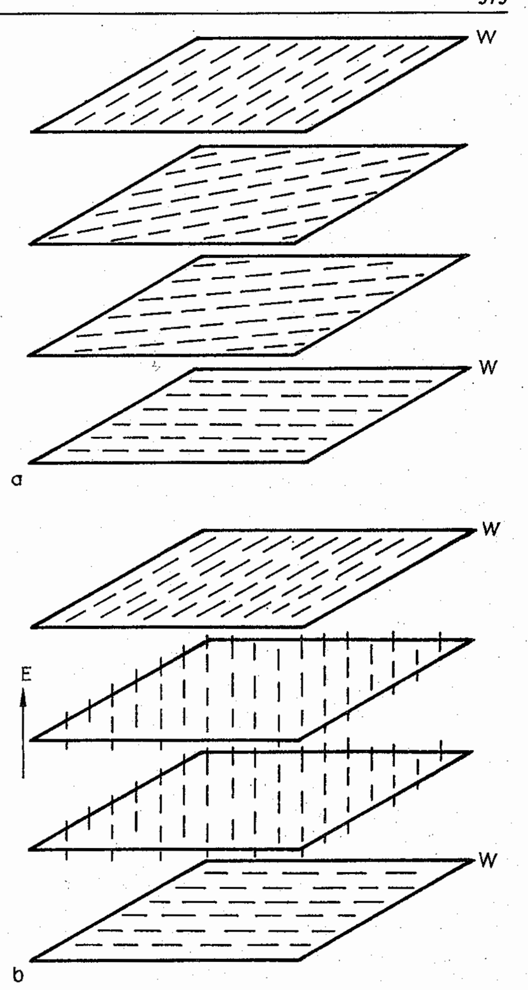


Abb. 8.1.6–19 Molekülanordnung in einem nematischen Flüssigkristall bei Steuerung der optischen Aktivität, (a) im feldfreien Fall und (b) bei anliegendem elektrischem Feld. Die zwischen den Gefäßwänden W gekennzeichneten Flächen sind willkürliche Schnitte, die die Moleküllagen im Volumen des Flüssigkristalles darstellen.

Gefäßwänden (vgl. Abb. 8.1.6–19). Läßt man linear polarisiertes Licht auf eine Gefäßwand fallen, so sind feldfreie Gebiete von Gebieten mit wirksamem Feldeffekt durch ihre Sichtbarkeit voneinander deutlich unterscheidbar.

In Flüssigkristall-Anzeigeelementen (Abb. 8.1.6–20) sind die Gefäßwände, die mit Siliciumdioxid (SiO_2) gegen die Substanz des Flüssigkristalls isoliert sind, mit den Segment- und Rückelektroden der darzustellenden Zeichen versehen.

Den Glasplatten, die das Flüssigkristallvolumen einschließen, folgen beiderseits Polarisatoren. Den Abschluß bildet ein Reflektor. Die Polarisationsfilter vor und hinter der Flüssigkristallschicht besitzen Polarisationsebenen, die um 90° zueinander gedreht sind. Die ebenfalls um 90° verdrillte nematische Struktur des Flüssigkristalles bewirkt im feld-

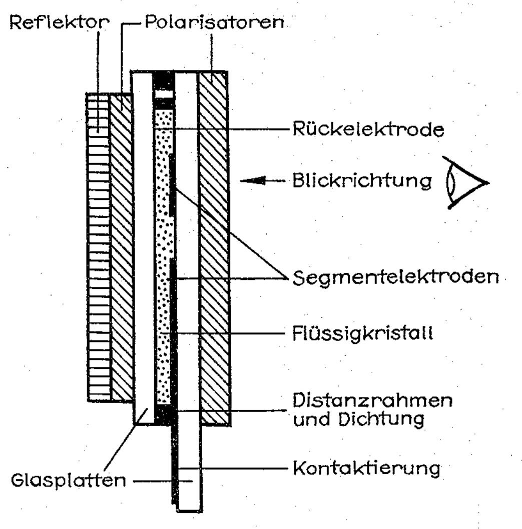


Abb. 8.1.6–20 Aufbau eines Flüssigkristallanzeigeelementes

freien Fall gerade diese 90°-Drehung der Polarisationsebene, so daß der Betrachter Lichtdurchlässigkeit und damit eine helle Fläche wahrnimmt. Bei angelegtem Feld stellen sich die Flüssigkristallmoleküle im Bereich der (durchsichtigen) Segmentelektroden parallel zur Blickrichtung und die Anordnung wird wegen der dadurch verlorengegangenen optischen Aktivität in dem Gebiet, in dem der Feldeffekt wirksam ist, undurchsichtig und erscheint dem Betrachter dunkel. Die außerordentlich geringe erforderliche Leistung (wenige Mikrowatt je Quadratzentimeter) führte zur völligen Verdrängung von Lumineszenzdioden für batteriegetriebene Taschenrechner und elektronische Uhren. Jedoch sind Flüssigkristallanzeigeelemente nur bei auffallendem Licht ablesbar. Einer weiteren Anwendung für Anzeigezwecke (z.B. Fernsehbildschirm) steht vorläufig die relativ große Trägheit, mit der die Molekülumorientierung unter dem Einfluß eines elektrischen Feldes erfolgt, im Wege. Die Einstellzeiten liegen bei 100 ms. Flüssigkristall-Anzeigeelemente werden mit Spannungsimpulsen angesteuert, um Elektrolysevorgänge zu vermeiden. Eine Ziffer kann durch sieben Segmente dargestellt werden. Einschließlich des Dezimalpunkts sind damit für die Ansteuerung acht Zuleitungen je Ziffer erforderlich. Die große Anzahl von Zuleitungen, die z. B. eine Taschenrechneranzeige nötig macht, ist technologisch durch die Vielzahl der Anschlüsse und durch den großen Platzbedarf nicht realisierbar. Eine Reduzierung der Anschlüsse ergibt sich durch den Multiplexbetrieb, bei dem mehrere Segment- bzw. Rückelektroden miteinander verbunden sind. Die Ansteuerung der einzelnen Segmente setzt dann ein speziell festgelegtes Impulsschema voraus.

8.1.6.10. Integrierte Optik stellt sich als Konsequenz und Erweiterung der Technologie integrierter Halbleiterschaltungen auf optische Systeme bei Anwendung optoelektronischer und optischer Wirkprinzipien dar. Es wird angestrebt, die Vorteile, die eine Signalübertragung auf optischem Wege besitzt, mit Hilfe der gut beherrschten technologischen Prinzipien der Mikroelektronik (z. B. der lithografischen Verfahren) handhabbar zu machen und auch hier von den diskreten optoelektronischen Bauelementen (z. B. Halbleiterlaser, LED, Fotodiode) zu integrierten, möglichst monolithischen Strukturen überzugehen. Diese erst in Ansätzen verwirklichten Vorstellungen erfordern bei Nutzung von meist bekannten physikalischen Effekten ein neuartiges Herangehen an die Probleme der Signalverarbeitung. Anstelle des metallischen Leiters tritt eine lichtleitende Schicht, die z.B. durch Katodenzerstäubung auf einem Trägerglas mit niedrigem Berechnungsindex hergestellt werden kann. Durch Ionenimplantation läßt sich in einem Träger ein Streifenleiter erzeugen, der auch für eine seitliche Begrenzung des Lichtweges sorgt. Neben der Fortleitung des Lichtes sind die mit der Einkopplung und der Beeinflussung der Lichtwellen verbundenen Fragen zu lösen. Solange Lichtquelle (z.B. GaAs-Laser) und Streifenleiter (z.B. BaO-SiO_2) aus verschiedenen Grundmaterialien bestehen, ist eine Einkopplung des Lichtes über ein Prisma oder eine Gitterstruktur möglich. Jedoch wird an Dünnschichtlaserstrukturen in Gelatineschichten gearbeitet, die mit ultraviolettem Licht gepumpt werden können und in Schichtsysteme integrierbar sind. Für die Verzweigung von Lichtwellen besteht die Möglichkeit der Strahlteilung durch ein diagonal am Kreuzungspunkt der Lichtwege stehendes Plättchen mit erhöhtem Brechungsindex. Da typische Längen integrierter optischer Systeme von der Größenordnung

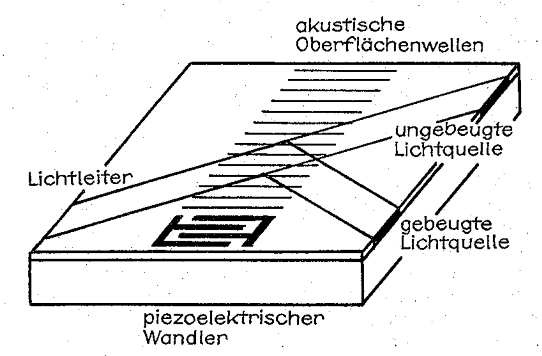


Abb. 8.1.6–21 Weiche für Lichtwellen infolge Beugung an akustischen Oberflächenwellen (schematisch)

der Lichtwellenlänge sind, ist zur theoretischen Behandlung die geometrische Optik nicht mehr adäquat und es muß auf die Wellenoptik zurückgegriffen werden. Eine interessante Lösung für eine Weiche ist ein akustooptischer Wandler (Abb. 8.1.6–21). Durch Piezoelektrische Wandler (Interdigitalwandler) werden in einem Träger akustische Oberflächenwellen erzeugt, die für die Lichtwellen eine Gitterstruktur darstellen, an der dann eine Beugung erfolgt. Die Ausbreitungsrichtung der Lichtwellen erfolgt, je nachdem, ob Oberflächenwellen erzeugt werden oder nicht, in die abgebeugte oder in die ursprüngliche Richtung (vgl. 8.1.5.).

8.1.7. Thermoelektrische Bauelemente

8.1.7.1. Thermoelemente und thermoelektrische Generatoren. Ströme von Ladungsträgern transportieren sowohl elektrische Ladung wie auch Energie. Durch einen im Halbleiter vorhandenen Temperaturgradienten wird das

Verhalten der Ladungsträger beeinflußt. Ihre Konzentration wird ortsabhängig, es entsteht ein Konzentrationsgradient. Gradienten der Ladungsträgerkonzentration haben Diffusionsströme (vgl. 8.1.3.1.) zur Folge, die im Sinne des Abbaus der Konzentrationsunterschiede fließen. Erweist sich weiterhin die Lage des Leitungsbandrandes als temperaturabhängig und damit bei Vorhandensein eines Temperaturgradienten als ortsabhängig, kommt ein Elektronenstrom in Richtung geringerer potentieller Energie der Ladungsträger zustande. Schließlich weisen Ladungsträger in Gebieten erhöhter Temperatur eine größere mittlere thermische Geschwindigkeit auf. Für die diffundierenden Ladungsträger stellt sich eine Vorzugsrichtung aus Bereichen höherer Temperatur in Bereiche tieferer Temperatur ein.

Erscheinungen, die bei gleichzeitiger Existenz eines elektrischen Feldes E und einer Temperaturdifferenz ΔT auftreten, bezeichnet man als thermoelektrische Effekte.

Mit jedem elektrischen Strom ist auch ein Wärmestrom verbunden, der in Richtung des Stromes der Ladungsträger fließt. Der Wärmestrom hängt mit der von den Ladungsträgern mitgeführten Energie zusammen. In einem Halbleiter mit einer stationären Temperaturdifferenz ΔT fließt also ein Elektronenstrom vom Gebiet höherer zum Gebiet tieferer Temperatur und transportiert Wärmeenergie. Im stromlosen Fall führt das im Halbleiter zu Oberflächenladungen. Diese Ladungen erzeugen ihrerseits ein elektrisches Feld, das Ursache für einen Strom ist, der dem durch ΔT verursachten entgegengerichtet ist. Daß trotzdem Wärme vom Gebiet höherer auf das Gebiet tieferer Temperatur übergeht, liegt daran, daß der aus dem Gebiet höherer Temperatur kommende Strom die größere Wärmeenergie mitführt.

Die Ausbildung eines elektrischen Feldes in einem stromlosen Halbleiterstab, zwischen dessen Enden eine Temperaturdifferenz besteht, bezeichnet man als *Seebeck-Effekt*. Die den Seebeck-Effekt charakterisierende Materialgröße ist die differentielle Thermospannung ε, die die Thermospannungsänderung, bezogen auf das Temperaturintervall angibt.

Bringt man zwei aus verschiedenen Materialien A und B zusammengesetzte Leiter miteinander in Kontakt und hält beide Kontaktstellen auf verschiedenen Temperaturen $T_1 < T_2$, so kann man, wenn der Leiterkreis in einem Material aufgetrennt ist, eine Thermospannung U_{Therm} messen (Abb. 8.1.7–1a), die von den differentiellen Thermospannungen ε_A und ε_B der beiden Materialien abhängt

$$U_{\text{Therm}} = \int_{T_1}^{T_2} (\varepsilon_A - \varepsilon_B)\,\mathrm{d}T.$$

Diesem Prinzip liegen die zur Temperaturmessung auf elektrischem Wege benutzten Thermoelemente zugrunde. Für häufig verwendete Materialpaare, z.B. Kupfer-Konstantan ($\varepsilon = 4{,}2 \cdot 10^{-5}\,\mathrm{VK}^{-1}$) und Platin-Platinrhodium ($\varepsilon = 6 \cdot 10^{-6}\,\mathrm{VK}^{-1}$) ist in nicht zu großen Temperaturintervallen die Thermospannung der Temperaturdifferenz proportional. In Halbleitern liegen die differentiellen Thermospannungen ε mehrere Größenordnungen höher und erreichen Werte von $10^{-3}\,\mathrm{VK}^{-1}$. Das erschließt neue Anwendungen. So läßt sich mit der experimentellen Anordnung eines Thermoelements nicht nur eine Temperaturmessung auf elektrischem Wege durchführen, sondern auch für spezielle Zwecke aus einem primär gegebenen Temperaturgefälle die Direktumwandlung von thermischer in elektrische Energie betreiben und somit ein thermoelektrischer Generator (Abb. 8.1.7–1b) aufbauen. Ein Wirkungsgrad von

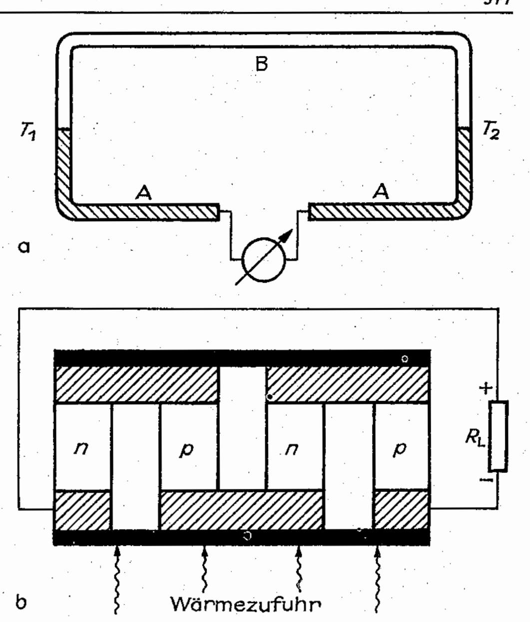


Abb. 8.1.7–1 Thermoelektrischer Effekt, (a) Thermoelement, (b) thermoelektrischer Generator (schematisch)

weniger als 10 % beschränkt den Einsatz jedoch auf Sonderfälle.

Als Substanzen, die eine hohe differentielle Thermospannung ε, eine große elektrische Leitfähigkeit σ und eine möglichst kleine Wärmeleitfähigkeit $\varkappa$ besitzen, wie sie für einen thermoelektrischen Generator zu fordern sind, eignen sich Verbindungen aus Bismut oder Blei mit Tellur.

8.1.7.2. Halbleiterkühlelemente. Der Seebeck-Effekt ist umkehrbar. Führt man einer einem Thermoelement entsprechenden Anordnung eine Spannung zu, so fließt ein Strom, und es tritt zwischen den beiden Kontaktstellen, an denen die zwei Substanzen verbunden sind, eine Temperaturdifferenz (Abb. 8.1.7–2) auf. Diese Erscheinung wird als Peltier-Effekt bezeichnet. Der an den Kontaktstellen erzeugte bzw. vernichtete Wärmestrom $\Delta Q/\Delta t = w$ ist der Stromdichte j proportional und wechselt mit ihr das Vorzeichen $w = \Pi j$.

Der Proportionalitätsfaktor Π ist der Peltier-Koeffizient. Er

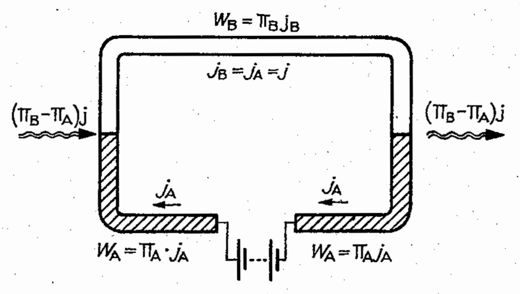


Abb. 8.1.7–2 Peltier-Effekt und Halbleiterkühlelement (schematisch)

beschreibt die im isothermen Fall von einem elektrischen Strom mitgeführte Wärme. Der Peltier-Koeffizient ist mit dem Seebeck-Koeffizienten durch die Beziehung $\Pi = \varepsilon T$ verbunden. Der Peltier-Koeffizient hat in *n*- und *p*-leitendem Material unterschiedliche Vorzeichen, verschwindet also in eigenleitenden Halbleitermaterialien.
Für im Kontakt befindliche Materialien mit unterschiedlichen Peltier-Koeffizienten (Abb. 8.1.7–2) $\Pi_A < \Pi_B$ wird Wärme von der linken zur rechten Kontaktstelle transportiert. Die Abkühlung der linken Kontaktstelle wird in *Halbleiterkühlelementen (Peltier-Elementen)* praktisch genutzt, wenn ihr Wärmeaustausch mit der Umgebung unterbunden wird.
Halbleiterkühlelemente aus Bismuttellurid (Bi_2Te_3) benötigen Gleichströme von einigen 10 A und wenigen Volt Spannungen. Bei guter thermischer Isolation werden Temperaturdifferenzen bis zu 50 K erreicht. Ihr prinzipieller Aufbau entspricht dem der thermoelektrischen Generatoren. Anstelle des Verbrauchers R_L tritt (vgl. Abb. 8.1.7–1) die Spannungsquelle. Bei Umkehr der Stromrichtung heizen sich die vorher gekühlten Kontaktstellen auf.

8.2. Integrierte Festkörperschaltkreise

Die technischen Anordnungen zur Erzeugung, Verstärkung, Übertragung und Verarbeitung von elektrischen Signalen, d. h. von in elektrischer Form vorliegenden Informationen bestehen aus Bauelementen, Zusammenschaltungen von Bauelementen, Geräten und Gerätesystemen. Nach der Entwicklung der Bauelemente und ihrer Stellung innerhalb von Gerätesystemen können vier Bauelementegenerationen gesehen werden. Zur ersten Generation zählen die Elektronenröhren (vgl. 8.1.2.). Zusammen mit der zweiten Bauelementegeneration, den Transistoren (vgl. 8.1.3.3. und 8.1.3.4.) stellen sie diskrete Einzelbauelemente dar. In der dritten Generation tritt ein neues Moment in die Entwicklung ein, das Ursache für einschneidende strukturelle Veränderung in der elektrischen Übertragung von Informationen wurde. Das neue Moment ist die Bauelementeintegration. In der dritten Generation bilden die einbezogenen Bauelemente und die Schaltung, die sie verbindet, eine Einheit. Ein unmittelbarer Zugriff zum Einzelbauelement ist nicht mehr möglich. Durch Hinzunahme einer noch wachsenden Zahl von Einzelbauelementen, beispielsweise für einen Herzschrittmacher, eine elektronische Uhr oder einen Taschenrechner, entwickelte sich die Großintegration, die man als vierte Bauelementegeneration ansehen kann. Als stimulierende Kräfte für die Miniaturisierung wirkten die wachsenden zu verarbeitenden Informationsmengen, die Erhöhung der Lebensdauer des Bauelements und weitere ökonomische Gesichtspunkte, wie die Verbilligung durch Verwendung hoher Stückzahlen von standardisierten Schaltungen und die Senkung des spezifischen Energieverbrauchs. In dieser Entwicklung leistete die Physik wesentliche Beiträge, denen besonders Ergebnisse der Festkörper- und Halbleiterphysik zugrundeliegen. Jedoch sind an der Herausbildung der Mikroelektronik auch zahlreiche andere Wissenschaftsdisziplinen beteiligt, unter denen sich Chemie, Kristallografie, Werkstoffwissenschaften und eine Reihe technischer Disziplinen befinden.

8.2.1. Technologien der Miniaturisierung

In der durch Elektronenröhren vertretenen ersten Bauelementegeneration bestanden die Geräte aus einer metallischen Trägerplatte, dem Chassis, auf dem die Bauelemente durch eine dreidimensionale Verdrahtung zusammengeschaltet waren. Der Übergang zu Halbleiterbauelementen ließ die zweidimensionale „Verdrahtung" auf Leiterplatten zur typischen Methode für die Zusammenschaltung der Einzelbauelemente werden. Die Bauelementeverkleinerung erfolgte über verschiedene Entwicklungslinien, z. B. über die Mikromodultechnik mit ihren in den geometrischen Dimensionen reduzierten Bauelementen oder die Schichttechniken. Die Dickschichttechnik mit Schichtdicken $\gg 1\,\mu m$ nutzt das Siebdruckverfahren, um vorwiegend Widerstände und Kondensatoren durch Auftragen einer geeigneten Paste herzustellen. In der Dünnschichttechnik mit Schichtdicken $< 1\,\mu m$ werden die Bauelemente durch Verarbeitung von Metallen, polykristallinen und amorphen Halbleiterschichten auf isolierenden Trägern durch Bedampfen oder Zerstäuben im Vakuum hergestellt. Universell ist die Halbleiterblocktechnik verbreitet. Hier werden auf einer einkristallinen Siliciumscheibe von einigen Zentimetern Durchmesser und einer Dicke bis zu $\approx 0{,}5$ mm viele integrierte Schaltungen gleichzeitig hergestellt und vor der Montage in die Gehäuse in die Teile (Chips) vereinzelt. Die am häufigsten verwendeten technologischen Varianten gründen sich auf die zur Herstellung von Bipolar- bzw. Feldeffekttransistoren nötigen Verfahrensschritte. Zur Einordnung integrierter Festkörperschaltkreise wird der Integrationsgrad als Kenngröße benutzt.

8.2.2. Halbleiterblocktechnik

Gegenüber den Schichttechniken, bei denen die isolierende Unterlage (z. B. Gläser oder Saphir) nur eine mechanische Aufgabe hat, befindet sich bei der Halbleiterblocktechnik die gesamte Schaltung in einem Halbleitereinkristall, der die elektronische Funktion und die Funktion des mechanischen Trägers übernimmt. Die Herstellung integrierter Schaltkreise nach der Halbleiterblocktechnik erfolgt in vier Arbeitsetappen (Zyklen), die jedoch nicht vollständig unabhängig voneinander sind und auch nicht voll mit dem chronologischen Fertigungsablauf übereinstimmen. Dem eigentlichen Herstellungsprozeß vorausgehen muß der Schaltungsentwurf und die daraus folgende Anfertigung spezieller Schablonen. Die Erzeugung von definierten Strukturen auf der Halbleiteroberfläche, die aus isolierenden, halbleitenden oder metallischen Bereichen bestehen können, wird durch Abdeckung der für eine Behandlung nicht vorgesehenen Gebiete erreicht.
Der *Zyklus 0* enthält die zur Bereitstellung der Siliciumeinkristallscheiben nötigen Arbeitsschritte, zu denen die Züchtung hochreiner Siliciumeinkristalle, die Scheibenbereitstellung und die mechanische, chemische und physikalische Bearbeitung der Scheibenoberfläche gehören.
Das Ergebnis des Zyklus 0 sind Siliciumscheiben der Dicke 100 bis 200 µm, mit Oberflächenrauhigkeiten $\ll 1\,\mu m$ und kristallografisch baufehlerfreier (111)- oder (100)-orientierter Oberfläche.
Im *Zyklus I* werden im einkristallinen Halbleitermaterial seitlich und in der Tiefe begrenzte Bereiche mit vorgegebenen elektrischen Eigenschaften erzeugt, die dann Widerstände, Dioden, Transistoren und die für ihr Funktionieren

nötigen leitenden Verbindungen darstellen. Im Verlaufe des Zyklus I erfolgen eine Anzahl von Verfahrensschritten, von denen sich einige mehrfach wiederholen können. Zu diesen Verfahrensschritten gehören Epitaxie, Oxidation, Diffusion, Implantation und Metallbedampfung.
Unter *Epitaxie* versteht man das kristallografisch entsprechend der Unterlage orientierte Aufwachsen einer einkristallinen Schicht, die durch eine geeignete Dotierung vorgegebene, von der Unterlage abweichende elektrische Eigenschaften erhält. Die *Oxidation* erfolgt dadurch, daß Sauerstoff und Wasserdampf bei Temperaturen um 1150 °C über die Siliciumscheiben geleitet werden. Es entsteht eine Isolatorschicht aus Siliciumdioxid (SiO_2). Solche Isolatorschichten lassen sich als Maskierungsschicht für einen später erfolgenden Diffusionsschritt verwenden. Zunächst wird die Isolationsschicht mit Fotolack beschichtet. Nachdem die Fotoschablone, die die Strukturdetails des künftigen Schaltkreises enthält, aufgelegt, justiert und belichtet wurde, erfolgt die fotografische Entwicklung. An definierten, durch die Schablone vorgegebenen Stellen ist die Halbleiteroberfläche durch den Lack geschützt. Mittels einer flußsäurehaltigen Ätzlösung können nun an den lackfreien Stellen der SiO_2-Isolatorschicht Fenster geätzt werden. Nach dem Entfernen der Lackreste werden durch die Öffnungen die Dotierungselemente (z. B. Bor, Gallium bzw. Phosphor, Arsen) bei Temperaturen um 1000 °C eindiffundiert. Der beschriebene fotolithografische Prozeß und die Diffusion können sich in Abhängigkeit von der Schaltkreisstruktur mehrfach mit jeweils verschiedenen Schablonen wiederholen. Die Schablonen müssen sich in ihren Größen bis auf Zehntel Mikrometer decken und eine entsprechende Positioniergenauigkeit der aufeinanderfolgenden Schablonenlagen muß gewährleistet sein.
Für hochintegrierte Schaltkreise ist aufgrund der kleiner werdenden Strukturen eine präziser einstellbare Dotierung erforderlich. Zu diesem Zweck eignete sich die *Ionenimplantation.* Die Dotierungsatome werden in einer Ionenquelle ionisiert, durch elektrische Felder beschleunigt und in die Halbleiterfläche eingeschossen. Vorteilhaft sind die einstellbaren steilen Dotierungsprofile. Zur Herstellung der leitenden Verbindungen erfolgt im Hochvakuum eine *Bedampfung* mit Aluminium, nachdem an den Kontaktstellen eine die gesamte Oberfläche bedeckende Isolatorschicht auf fotolithografischem Wege entfernt wurde. Nach einem weiteren Überzug mit Fotolack wird durch eine Schablone dort belichtet, wo eine Leiterbahn entstehen soll. Nach der Entwicklung wird das Aluminium an den nichtbelichteten Stellen abgeätzt. Nach Ablauf dieser Verfahrensschritte liegt eine Siliciumscheibe vor, die mehrere hundert integrierte Schaltungen enthält. Es schließt sich eine Messung der Funktionstüchtigkeit der Schaltkreise auf der Scheibe an. Dabei werden Schaltkreise, die wesentliche Parameter nicht einhalten, gekennzeichnet.
Zum *Zyklus II* gehört das Vereinzeln der Schaltkreise. Zu diesem Zweck werden die Mittellinien zwischen den Schaltkreisrändern mit einem Diamanten geritzt bzw. mit einem Laserstrahl zum Brechen vorbereitet. Die kristallografische Orientierung wird so gewählt, daß die Bruchflächen mit den natürlichen (110)-Spaltflächen der Diamantstruktur des Siliciums übereinstimmen. Es schließt sich die Befestigung des Plättchens auf einem Trägerstreifen an, die Kontaktierung der Inseln des Schaltkreises erfolgt mit Gold- oder Aluminiumdrähten von etwa 20 µm Durchmesser durch Thermokompression, Schweißen oder Ultraschall. Nach dem Gehäuseverschluß, der den Zyklus II beendet, schließt sich *Zyklus III* an, der die Messung und Prüfung der fertigen Schaltkreise zum Inhalt hat.

8.2.2.1. Integrationsgrad. Zur Festlegung des Integrationsgrades, eines Charakteristikums integrierter Schaltkreise, wird die Anzahl der Funktionselemente je Chip benutzt. Funktionselemente in integrierten Schaltkreisen können Widerstände, Kondensatoren, *pn*-Übergänge oder Transistorstrukturen sein. Zur *Kleinintegration* zählen Schaltkreise bis zu 100 Funktionselementen, das können logische Gatter (vgl. 8.2.4.2.) oder einfache Verstärker sein. Schaltkreise der *Mittelintegration* enthalten 10^2 bis 10^3 Funktionselemente. Zu ihnen gehören in der Analogelektronik Operationsverstärker (vgl. 8.2.3.1.) und in der Digitalelektronik Addierer (vgl. 8.2.4.3.), Dekodierer (vgl. 8.2.4.3.) und Schieberegister (vgl. 8.2.4.4.). Von *Großintegration* spricht man bei Schaltkreisen mit 10^3 bis 10^4 Funktionselementen je Chip. Hier sind Analog-Digital-Wandler (vgl. 8.2.5.), Mikroprozessoren (8.2.6.4.), Speicher (vgl. 8.2.6.3.) und Taschenrechnerschaltkreise einzuordnen. In Schaltkreisen mit dem Integrationsgrad *Höchstintegration*, die zwischen 10^4 und 10^5 Funktionselemente enthalten, sind Speicher großer Kapazität und Mikrorechner zu realisieren.

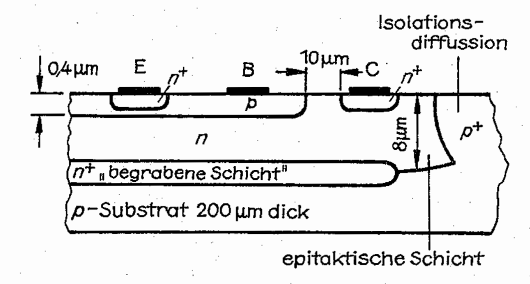


Abb. 8.2.2-1 Typische Größen eines integrierten *npn*-Transistors

8.2.2.2. Zur Technologie integrierter Bipolarschaltkreise. Die zahlreichen, von den Produzenten angewandten Varianten zur Herstellung von integrierten Schaltkreisen gehen auf zwei Basistechnologien zurück. Das eine dieser Grundverfahren ist das bei der Herstellung von Bipolartransistoren (vgl. 8.1.3.3.) beschriebene. Die für Bipolarschaltkreise zu erfüllende, neuhinzukommende Forderung, die im Volumen des Halbleiterkristalls enthaltenen Bauelemente gegeneinander elektrisch zu isolieren, läßt sich auf unterschiedliche Weise verwirklichen. Am einfachsten und ohne zusätzliche Prozeßschritte erfolgt eine Isolation durch in Sperrichtung vorgespannte *pn*-Übergänge (vgl. Abb. 8.2.2-1). Nachteilig für das Hochfrequenzverhalten solcher Schaltkreise sind jedoch die unerwünschten, aber nicht vermeidbaren parasitären Sperrschichtkapazitäten. Aufwendigere Isolationsverfahren sorgen für eine Trennung durch tiefe isolierende SiO_2-Gräben, die bis zum Substrat reichen, oder ätzen das Silicium in der Umgebung der Inseln ab, so daß die Verbindungsleitungen selbsttragend werden. Integrierte Bipolarschaltkreise werden sowohl in der Analogelektronik (vgl. 8.2.3.) wie auch in der Digitalelektronik (vgl. 8.2.4.) angewandt. Als bipolare analoge Schaltkreise gibt es u. a. verschiedenartige Verstärker- und Empfängerschaltungen, Zwischenfrequenzverstärker, Stereodecoder und spezielle Schaltkreise für Fernsehempfänger. Wegen seiner großen Anwendungsbreite in der Analogelek-

tronik hat der integrierte Operationsverstärker (vgl. 8.2.3.1.) besondere Bedeutung.
Bipolare Schaltkreise werden auch in der Digitaltechnik angewandt. Durch unterschiedliche Herstellungsverfahren ergeben sich Schaltkreisfamilien mit voneinander abweichenden Kennwerten.
Da es in der Digitaltechnik nur diskrete (meist zwei) Signalzustände gibt, ist die Schaltgeschwindigkeit bzw. die Verzögerungszeit eine solche Kenngröße. Den Kehrwert des Produkts von Verzögerungszeit und Verlustleistung eines logischen Grundgatters (typische Werte 10 bis 100 pWs) bezeichnet man auch als dessen Güte. Die am häufigsten anzutreffenden digitalen Bipolarschaltkreise gehören zur TTL-Familie (*T*ransistor-*T*ransistor-*L*ogik). Ihre Grundschaltung, die die logische Verknüpfung eines negierten UND darstellt und einen Transistor mit Mehrfachemitter

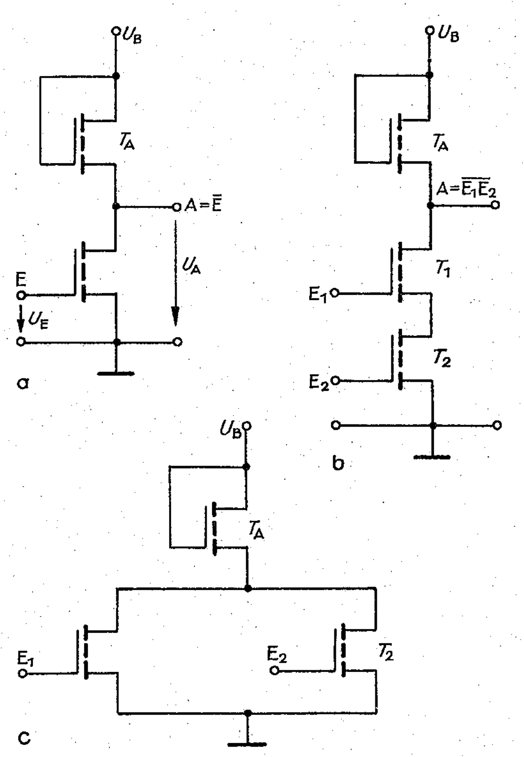


Abb. 8.2.2–2 Digitale Grundschaltungen mit Feldeffekttransistoren, (a) Negator, (b) NAND, (c) NOR (T_A = Arbeitswiderstand)

am Eingang aufweist, wird in 8.2.4.2. beschrieben. In der TTL-Technik wird der Transistor, der sich im Zustand „Schalter geschlossen“ befindet, im Sättigungsbereich betrieben. Er hat durch den beim Umschalten nötigen Ladungstransport eine größere Schaltzeit, was bei der Schaltkreisfamilie ECL (emittergekoppelte Logik), bei der der Transistor im aktiven Bereich arbeitet, nicht auftritt. Wachsende Bedeutung, besonders für großintegrierte Schaltkreise, gewinnt die integrierte Injektionslogik (IIL = I^2L), die für ein Grundgatter nur 10 % der Chipfläche eines TTL-Gatters benötigt. Verschiedene Beispiele für digitale Bipolarschaltkreise sind die Decodierer (vgl. 8.2.4.3.), Flipflops (vgl. 8.2.4.4.), Zähler (vgl. 8.2.4.4.) und Schieberegister (vgl. 8.2.4.4.).

8.2.2.3. Zur Technologie der integrierten MIS-Schaltungen. Die zweite wichtige Basistechnologie baut auf der Herstellung von Feldeffekttransistoren (vgl. 8.1.3.4.) auf. Die für bipolare integrierte Schaltkreise nötigen Isolationsverfahren zur elektrischen Trennung entfallen bei integrierten MIS-Schaltungen, da die MIS-Transistoren von ihrem Wirkungsprinzip her selbstisolierend sind. Sie erfordern daher wesentlich weniger Chipfläche und sind einfacher herstellbar. Ein MIS-Schaltkreis kann mit einem Diffusionszyklus auskommen, während bipolare Schaltkreise in der Epitaxie-Planar-Technologie meist vier Diffusionszyklen enthalten. Kennzeichnend für MIS-Schaltungen ist, daß auch Widerstände und Kondensatoren mit dem MIS-Feldeffekttransistor realisiert werden. Im Standardverfahren zur Herstellung eines *p*-Kanal-Feldeffekttransistors mit metallischer Gate-Elektrode wird zunächst auf das *n*-leitende Substrat eine SiO_2-Schicht oxidiert. Nach dem Öffnen der Fenster zur *p*-Diffusion für die Source- und Drain-Gebiete wird erneut mit einer dicken Schicht oxidiert. Anschließend erfolgt die Beseitigung des Oxides über dem künftigen Kanal. Dort wird eine dünne Oxidschicht definierter Dicke (≈0,1 µm) hergestellt. Die letzten Verfahrensschritte sind Öffnung der Kontaktfenster für Source und Drain und die Aluminiumbedampfung der drei Elektroden. Im Vergleich zur Bipolartechnik gibt es weniger Verfahrensschritte zur Herstellung. Da jeweils zwischen Source, Kanal und Drain einerseits und dem Substrat andererseits ein *pn*-Übergang vorhanden ist, der die Selbstisolation des Feldeffekttransistors im Chipvolumen gegenüber benachbarten Transistoren bewirkt, entfallen im Vergleich zur Bipolartechnologie beispielsweise der Diffusionsprozeß zur Isolation und der Epitaxieprozeß. Insgesamt reduziert sich die Zahl der Prozeßschritte auf etwa ein Drittel. Der Flächenbedarf für ein Grundgatter beträgt gegenüber der bipolaren TTL-Technologie weniger als 10 %. Durch die nahezu leistungslose Steuerung der Feldeffekttransistoren ist der Leistungsverbrauch gering, die Schaltgeschwindigkeit verringert sich jedoch ebenfalls. Ihr wird durch Verkürzung der Kanallänge entgegengewirkt (Kurzkanaltechnik bis ≈1 µm). Von den vielen technologischen Varianten sei erwähnt, daß die Verwendung von polykristallinem Silicium als Gate-Elektrode eine Herabsetzung der Schwellspannung (Gate-Source-Spannung beim Einsetzen des Drain-Stromes) bewirkt und einen teilweisen Wegfall der Maskenjustierung erlaubt. Die Senkung der Schwellspannung ermöglicht eine Erniedrigung der Versorgungsspannung und bringt damit den Vorteil der Zusammenschaltung mit bipolaren TTL-Schaltkreisen (TTL-Kompatibilität). Auf die eintretenden Vorteile bei gleichzeitiger Herstellung von *n*-Kanal und *p*-Kanal-Feldeffekttransistoren (CMOS-Technik) wurde in 8.1.3.4. hingewiesen.
Aufgrund ihrer Eigenschaften erfolgt der Einsatz integrierter MIS-Schaltkreise in der Digitalelektronik auf breiter Front. Beispiele sind logische Grundgatter, Flipflops, Zähler, Frequenzteiler und Speicher. Der hohe Eingangswiderstand von Feldeffekttransistoren bewirkt eine relativ langsame (1 bis 100 ms) Entladung von Leitungs- und Gate-Kanal-Kapazitäten. Das kann zur kurzzeitigen Ladungsspeicherung genutzt werden. Lädt man in Zeitabständen, die kurz gegen die Entladezeiten sind, taktgesteuert nach, so lassen sich auf diese Weise Speicher sehr hoher Packungsdichte leistungsarm realisieren. Diese „dynami-

sche“ Schaltungstechnik ist besonders für hochintegrierte Schaltkreise (vgl. 8.2.6.) von Bedeutung.
Weitere hochintegrierte MIS-Schaltkreise sind Zeichengeneratoren, Rechnerschaltkreise, Mikroprozessor-Schaltkreise, Uhrenschaltkreise u. a.
Grundschaltung für die digitale Anwendung von Feldeffekttransistoren ist der Negator. Er besteht aus zwei MISFET. T_A stellt den Arbeitswiderstand dar (Abb. 8.2.2–2a). Die logische Funktion NAND ist durch in Reihe geschaltete MISFETs zu verwirklichen (Abb. 8.2.2–2b), NOR durch parallelgeschaltete Transistoren (Abb. 8.2.2–2c).

8.2.3. Integrierte Analogschaltungen

Neben den speziellen in 8.2.2.2. angeführten Einsatzbereichen standardisierter bipolarer integrierter Schaltungen in Rundfunk-, Fernseh- und Nachrichtentechnik gibt es auch universeller anwendbare integrierte Analogschaltungen. Der wichtigste Vertreter der für vielfältige Funktionen einsetzbare integrierte bipolare Analogschaltkreis ist der Operationsverstärker.

8.2.3.1. Operationsverstärker. Die Bezeichnung Operationsverstärker entstand bereits vor der Existenz integrierter Technologien. Sie wurde auf Verstärkerbausteine in Analogrechnern angewandt, die zur elektronischen Nachbildung von Rechenoperationen dienten. Durch die Herstellung als integrierte Schaltung ist der Operationsverstärker auf sehr vielen Gebieten der analogen Signalverarbeitung auf ökonomisch günstige Weise verwendbar.

Eigenschaften des Operationsverstärkers. Er ist ein mehrstufiger Gleichspannungsverstärker, hat eine sehr große (einige 10^4) Spannungsverstärkung, verfügt über einen Differenzspannungseingang, einen hohen Eingangs- und einen niedrigen Ausgangswiderstand. Aufgrund der sehr hohen Spannungsverstärkung hängt das Übertragungsverhalten des Operationsverstärkers praktisch nur vom Rückkopplungsnetzwerk zwischen Verstärkerausgang und -eingang ab.

Tab. 8.2.3–1 Kennwerte von idealen und realen Operationsverstärkern

Kennwert	Symbol	idealer OV	realer OV
Differenzspannungsverstärkung	v_D	∞	$10^4 \ldots 10^5$
Differenzeingangswiderstand	r_D	∞	200 kΩ
Ausgangswiderstand	r_A	0	150 Ω
Verstärkungsbandbreiteprodukt	$v_D \cdot b$	∞	1 MHz
Offsetspannung (u_D für $u_A = 0$)	u_{E_0}	0	1 mV

Schaltungen mit Operationsverstärkern sind auch ohne Kenntnis ihrer konkreten Innenschaltung berechenbar, wenn man einen idealen Operationsverstärker zugrunde legt. Tab. 8.2.3–1 stellt Kennwerte von idealen und realen Operationsverstärkern einander gegenüber. Abb. 8.2.3–2a zeigt das Schaltsymbol. Von den beiden Eingängen wird der mit einem + versehene als nichtinvertierend, der mit einem − versehene als invertierend gekennzeichnet. Abb. 8.2.3–2b gibt die Übertragungskennlinien wieder. Wenn die Differenzeingangsspannung null ist, liegt am Ausgang eines realen Operationsverstärkers eine kleine, von null verschiedene Ausgangsspannung, die herstellungsbedingte Ursachen hat. Die Eingangsspannung, die am Ausgang eine verschwindende Spannung bewirkt, bezeichnet man als Offsetspannung. Die Differenzspannungsverstärkung v_D ist, wie am Schaltsymbol erkennbar, $v_D = u_A/u_D = u_A/(u_P - u_N)$. Für $u_N = 0$ arbeitet der Operationsverstärker als nichtinvertierender, für $u_D = 0$ als invertierender, d. h. die Phase zwischen Eingangs- und Ausgangssignal um 180° drehender Verstärker. Die Verstärkung der beiden Grundschaltungen (Abb. 8.2.3–3) läßt sich für ideale Operationsverstärker und bei Betrachtung der Anordnungen als Spannungsteiler sofort hinschreiben. Die Spannungsverstärkung der invertierenden Verstärkerschaltung eines idealen

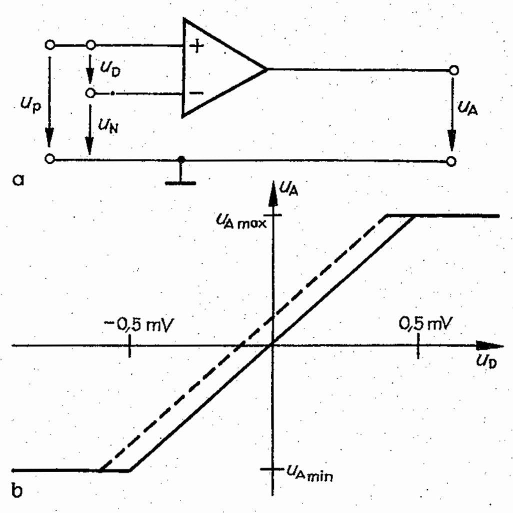


Abb. 8.2.3–2 Operationsverstärker, (a) Schaltsymbol, (b) Übertragungskennlinie (—— idealer, – – – realer Operationsverstärker)

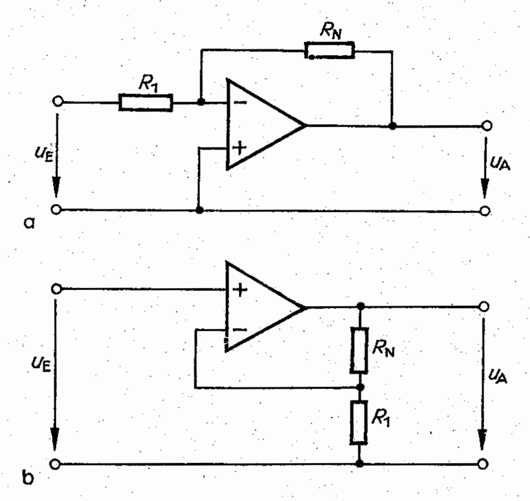


Abb. 8.2.3–3 Grundschaltungen von Operationsverstärkern, (a) invertierender Verstärker, (b) nichtinvertierender Verstärker

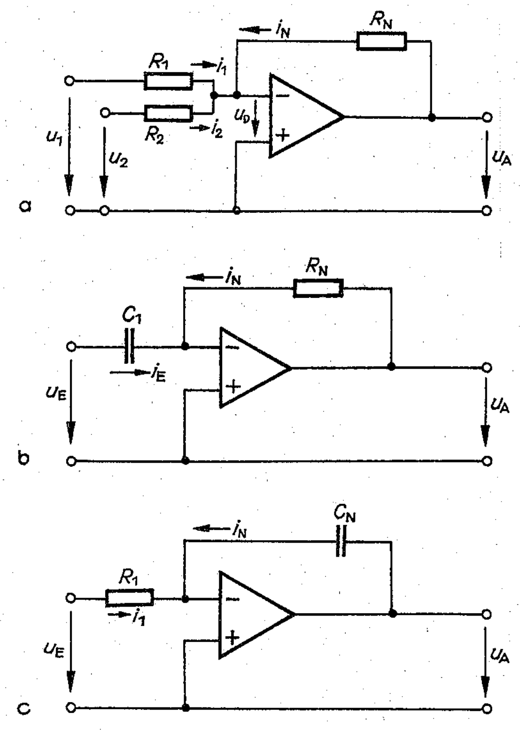


Abb. 8.2.3–4 Anwendungen des Operationsverstärkers, (a) Addierschaltung, (b) Differentiator, (c) Integrator

Operationsverstärkers (Abb. 8.2.3–3a) ist $v_u = -R_N/R_1$, wobei das Minuszeichen die Phasenverschiebung zwischen Ausgangs- und Eingangsspannung beschreibt. Die Verstärkung wird nur durch die äußere Beschaltung bestimmt. Beim realen Operationsverstärker bedingt die endliche Leerlaufverstärkung eine Korrektur. Die Spannungsverstärkung des nichtinvertierenden idealen Operationsverstärkers (Abb. 8.2.3–3b) ist $v_u = 1 + R_N/R_1$.

Entwicklungstendenzen bei integrierten Operationsverstärkern zielen auf eine einzige Betriebsspannung. Häufig müssen zur Stromversorgung noch eine positive und eine negative Betriebsspannung bereit gestellt werden. Vorteilhaft sind auch Kombinationen der Bipolar- und Feldeffekttransistor-Technologie (BIFET) für Operationsverstärker mit sehr hohen Eingangswiderständen ($> 10^{12}\,\Omega$).

Anwendungen des Operationsverstärkers. Als Beispiele seien eine Additionsschaltung für analoge Größen, eine Integrations- und eine Differentiationsschaltung angeführt (Abb. 8.2.3–4).

Für die Additionsschaltung gilt am invertierenden Eingang nach dem Knotensatz für den idealen Operationsverstärker ($v_D \to \infty$, $u_D \approx 0$)

$$i_1 + i_2 = -i_N$$

d. h. $$u_1/R_1 + u_2/R_2 = -u_A/R_N$$

oder $$(R_N/R_1)u_1 + (R_N/R_2)u_2 = -u_A$$

Falls $R_1 = R_2 = R_N$ wird $u_A = -(u_1 + u_2)$.

Durch einen nachgeschalteten invertierenden Verstärker der Spannungsverstärkung $v_u = 1$ wird das Vorzeichen der Ausgangsspannung umgekehrt, so daß insgesamt die Addition zweier analoger Größen ausgeführt wird.

Für die Differentiationsschaltung Abb. 8.2.3–4b gilt für den idealen Operationsverstärker $i_E = C_1\,du_E/dt = -i_N = -u_A/R_N$ oder $u_A = -C_1 R_N\,du_E/dt$.

Bei der Integrationsschaltung Abb. 8.2.3–4c hat man unter den gleichen Bedingungen $i_1 = u_E/R_1 = -i_N = -C_N\,du_A/dt$ oder $u_A = -1/(R_1C_N) \int u_E\,dt$.

Für die sehr vielseitige Anwendung des Operationsverstärkers in Abhängigkeit von seiner Außenbeschaltung seien weitere Beispiele aufgezählt. Durch geeignetes Einfügen von Kondensatoren in das Rückkopplungsnetzwerk entstehen Verstärker mit frequenzabhängiger Spannungsverstärkung, was Hoch- bzw. Tiefpaßschaltungen entspricht, wie sie bei Klangregel- und Entzerrerschaltungen eingesetzt werden können. Operationsverstärker können auch als Sinusgeneratoren, Impulsgeneratoren, als Komparatoren (Spannungsvergleicher), Konstantspannungs- und Konstantstromquellen u. a. arbeiten.

8.2.4. Digitale Schaltungen

Bei der Anwendung elektronischer Bauelemente unterscheidet man zwei Grundtypen von Schaltungen, die analogen und die digitalen. Sie unterscheiden sich durch die Struktur der informationstragenden Signale, die sie als Eingangsgrößen verarbeiten können. Die Bedeutung von „analog" (griech., gleichartig) ist in diesem Zusammenhang nur ungenau dem herkömmlichen Gebrauch dieses Begriffs zu entnehmen. Gemeint ist, daß analoge Schaltungen stetige, lückenlos veränderliche, d. h. über einen unendlichen Wertevorrat verfügende Eingangsgrößen verarbeiten. Als digital (lat., digitus, Finger) werden Schaltungen für diskret gestufte Eingangsgrößen bezeichnet. Die Übergänge zwischen den diskreten Signalzuständen erfolgen sprunghaft. Die Signalzustände sind durch verbotene Bereiche voneinander getrennt. Schaltungen für Signale, die nur zweier Zustände fähig sind (binäre Digitalschaltungen), haben sich aufgrund der großen Sicherheit gegen Störungen durchgesetzt. Diesen beiden Signalgrößen können die logischen Aussagen „wahr" und „falsch" oder auch die Ziffern des dualen Zahlensystems „L" und „Ø" zugeordnet werden. Die Entwicklung von digitalen Schaltungen hat mit dem Ausbau der Digitalrechentechnik (im Gegensatz zur Analogrechentechnik) besonderen Auftrieb erhalten.

Digitale Schaltungen erschließen sich zunehmend neue Anwendungsbereiche. Als Beispiele können Meßtechnik, Informationsübertragung, -speicherung und -verarbeitung und Unterhaltungselektronik dienen. Da viele Signale in analoger Form vorliegen, sind zur Transformation in eine digitale Form Analog-Digital-Umsetzer (vgl. 8.2.5.) erforderlich. Die elektronische Realisierung von Digitalschaltung zeichnet sich durch einige Besonderheiten aus. Für die in der Digitaltechnik einzusetzenden Bauelemente gilt im Vergleich zur Analogtechnik, daß die zugelassenen Toleranzen größer sein dürfen und daß der erforderliche Werteumfang des Bauelementesortiments wesentlich geringer ist. Zudem ist die Typenzahl von Grundschaltungen relativ klein, während die Stückzahl der Einzelschaltungen je Gerät relativ groß ist. In der Analogtechnik sind diese Tendenzen gerade umgekehrt. Als wesentliche Vorteile der Digitaltechnik können die durch Vergrößerung der Stellenzahl erreichbare hohe Genauigkeit, die Möglichkeit einer

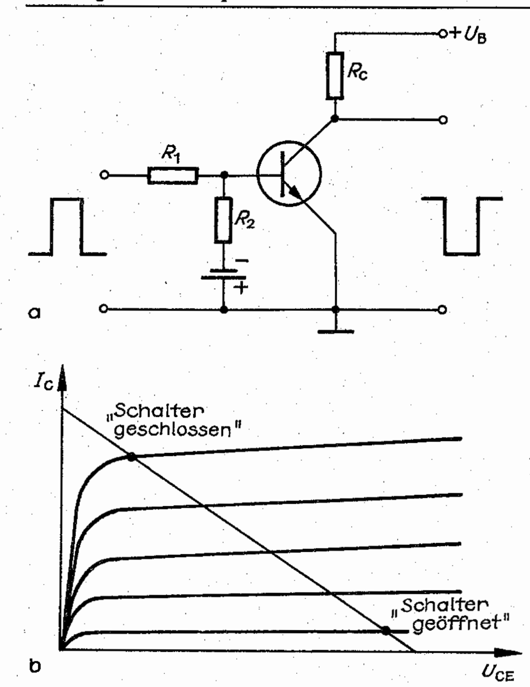


Abb. 8.2.4–1 Transistor als Schalter, (a) Schaltstufe, (b) Ausgangskennlinienfeld mit Arbeitspunkten

Signalregenerierung und damit die hohe Störsicherheit angesehen werden.

Die beiden Zustände lassen sich mit Hilfe von Transistoren herbeiführen, die als elektronische Schalter betrieben werden. Abb. 8.2.4–1 zeigt das am Beispiel eines Bipolartransistors. Der „Aus"-Zustand entspricht dem Sperrbereich des Transistors, der durch $U_{BE} < 0$ und $U_{BC} < 0$, also durch Sperrpolung von Basis-Emitter- und Basis-Kollektorübergang gekennzeichnet ist. Über den Transistor fließt nur der geringe Kollektorsperrstrom. Der Transistor wirkt wie ein geöffneter Schalter, an ihm fällt nahezu die gesamte Betriebsspannung U_B ab. Dem „Ein"-Zustand ist der Sättigungsbereich mit der Flußpolung beider *pn*-Übergänge mit $U_{BE} > 0$ und $U_{BC} > 0$ zugeordnet. Durch den Transistor fließt ein großer Kollektorstrom, der Spannungsabfall U_{CE} am Transistor ist klein, der Transistor entspricht in diesem Zustand einem geschlossenen Schalter.

Zur Interpretation der in binären Digitalschaltungen möglichen Verknüpfungen bedarf es eines mathematischen Hilfsmittels, der Schaltalgebra. Ihr liegt die Boolesche Algebra zugrunde, in der Zusammenhänge in einer zweiwertigen Aussagelogik formuliert sind.

8.2.4.1. Logische Grundfunktionen und ihre elektronische Realisierung. Die wichtigsten logischen Grundfunktionen mit ihren Wahrheitstabellen und Schaltsymbolen gibt Abb. 8.2.4–2 an. Sie enthält eine Spalte mit einer Kontaktanordnung. Da Relaiskontakte zweier Zustände fähig sind, sind hier die Kontaktstellungen als Signalgrößen in Kontaktnetzwerken den logischen Grundfunktionen zugeordnet. Man unterscheidet zwei Arten von Relaiskontakten, Arbeitskontakte und Ruhekontakte. Arbeitskontakte sind in Arbeitsstellung und Ruhekontakte in Ruhestellung geschlossen. Der Signalgröße „Kontakt in Arbeitsstellung" sei die Eins des Dualsystems bzw. die logische Aussage „wahr" (die ebenfalls durch „L" symbolisiert werde) zugeordnet. Der Signalgröße „Kontakt in Ruhestellung" entspricht die „0" bzw. die logische Aussage „falsch". So ist z. B. Verknüpfung zweier binärer Signale durch die Boolesche Funktion UND durch die Reihenschaltung zweier Arbeitskontakte charakterisiert. Die Negation stellt sich als Ruhekontakt dar. Eine andere, elektronische Realisierung der Negation enthält Abb. 8.2.4–1. In der angegebenen Transistorschaltung ist das Ausgangssignal zum Eingangssignal invers. Sie stellt also einen mit Hilfe eines im Schalterbetrieb arbeitenden Bipolartransistors realisierten Negator dar. Die logischen Grundfunktionen sind bei Nutzung verschiedener Bauelemente schaltungstechnisch auf vielfältige Weise zu realisieren. Auf die Erzeugung des UND- und ODER-Verhaltens mit Hilfe von Dioden ist z. B. in 8.1.3.2. verwiesen. Von praktischer Bedeutung sind ausnahmslos Systeme integrierter Digitalschaltkreise, die man je nach wesentlichen Realisierungsgesichtspunkten entsprechenden „Schaltkreisfamilien" zuordnet. Die zitierte Dioden-Transistor-Logik (DTL) hat an Bedeutung verloren. Am häufigsten verwendet wird die Transistor-Transistor-Logik (TTL). Die mittlere Verzögerungszeit, die vergeht, bis nach der Ansteuerung der Ausgang den neuen Zustand angenommen hat, beträgt hier etwa 10 ns und die Verlustleistung für eine logische Funktion liegt für Impulsfrequenzen von 20 MHz etwa bei 10 mW. Durch technologische Varianten lassen sich Schaltgeschwindigkeit und Verlustleistung beeinflussen. So erreicht die ECL-Familie (emittergekoppelte Logik mit Bipolartransistoren) Durchlaufverzögerungen je Grundfunktion von 0,8 bis 2 ns und typische Impulsfrequenzen von 150 MHz. Das muß jedoch mit einer höheren Leistungsaufnahme zwischen 25 und 60 mW erkauft werden. Die Schaltkreisfamilie mit unipolaren Metall-Oxid-Silicium-Transistoren (CMOS, vgl. 8.1.3.5.) zeichnet sich durch wesentlich reduzierte Verlustleistung und eine gegenüber Bipolartransistoren um das 5- bis 10fache höhere Packungsdichte aus. Die typischen Zeiten für die Durchlaufverzögerung, die die Rechengeschwindigkeit bestimmen, liegt jedoch bei 25 ns.

8.2.4.2. TTL-Grundschaltung. In der TTL-Technik stellt

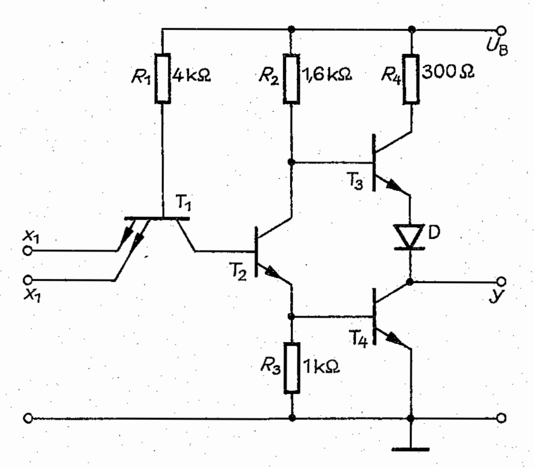


Abb. 8.2.4–3 Elektronische Realisierung der negierten Konjunktion (NAND) in der TTL-Standard-Schaltung

Boolesche Funktion	Mathem. Darstellung	Kontaktanordnung	Wertetabelle	Symbol
Konjunktion (UND)	$x_1 \wedge x_2 = y$	x_1 x_2	$x_2 \backslash x_1$: 0, L; 0: 0, 0; L: 0, L	&
Disjunktion (ODER)	$x_1 \vee x_2 = y$	x_1 x_2	$x_1 \backslash x_2$: 0, L; 0: 0, L; L: L, L	1
Negation (NICHT)	$\bar{x} = y$		x: 0, L; y: L, 0	
Negiertes UND (NAND)	$\overline{x_1 \wedge x_2} = y$	x_1 x_2	$x_2 \backslash x_1$: 0, L; 0: L, L; L: L, 0	&
Negiertes ODER (NOR)	$\overline{x_1 \vee x_2} = y$	x_1 x_2	$x_2 \backslash x_1$: 0, L; 0: L, 0; L: 0, 0	1
Äquivalenz (Exklusiv-NOR)	$(\bar{x}_1 \wedge \bar{x}_2) \vee (x_1 \wedge x_2) = y$	x_1 x_2	$x_2 \backslash x_1$: 0, L; 0: L, 0; L: 0, L	=
Antivalenz (Exklusiv-ODER)	$(\bar{x}_1 \wedge x_2) \vee (x_1 \wedge \bar{x}_2) = y$	x_1 x_2	$x_2 \backslash x_1$: 0, L; 0: 0, L; L: L, 0	=1

Abb. 8.2.4–2 Logische Grundfunktionen

die negierte Konjunktion, die NAND-Funktion, die integrierte Grundschaltung dar (Abb. 8.2.4–3). Das elektronische Verhalten bestimmt die Eingangsstufe mit dem Mehremitter-Transistor T_1 und dem Transistor T_2. Nimmt einer der beiden Emitter (z.B. x_1) niedriges Potential („Ø“) an, so ist der Basis-Emitter-*pn*-Übergang von T_1 in Flußrichtung gepolt. Basis und Kollektor von T_1 nehmen dann ebenfalls das niedrige Potential an. Das führt auch an der Basis von T_2 zu einem niedrigen Potential, so daß T_2 gesperrt ist. An R_2 tritt somit kein Spannungsabfall auf, und der Kollektor von T_2 liegt auf hohem Potential „L“). Der auf hohem Potential befindliche Emitter x_2 hat auf das Ergebnis keinen Einfluß, da sein Basis-Emitter-*pn*-Übergang in Sperrichtung gepolt ist. Befinden sich x_1 und x_2 gleichzeitig auf niedrigem Potential („Ø“), bleibt das Kollektorpotential von T_2 hoch („L“). Liegen beide Emitter gleichzeitig auf hohem Potential („L“), so sind die Basis-Emitter-Übergänge von T_1 gesperrt. Dafür fließt über die in Flußrichtung gepolte Basis-Kollektor-Strecke von T_1 ein Strom, der von der ebenfalls in Flußrichtung gepolten Basis-Emitter-Strecke des Transistors T_2 zugeführt wird. T_1 wird hierbei im aktiv-inversen Zustand (Funktion von Emitter und Kollektor vertauscht) betrieben. Damit befindet sich T_2 im leitenden Zustand und das Kollektorpotential von T_2 ist gering („Ø“). Insgesamt wird durch die vier möglichen Belegungen an x_1 und x_2 durch die Eingangsstufe mit T_1 und T_2 die logische Verknüpfung NAND erzielt. Der Mehremitter-Transistor T_1 bewirkt dabei die Konjunktion und T_2 die Negation. Für die praktische Anwendung ist die Zuschaltung einer leistungsfähigen Ausgangsstufe nötig, die den Anschluß weiterer logischer Schaltungen an den Ausgang erlaubt und die Spannungsdifferenz zwischen den beiden Schaltzuständen hinreichend groß hält. Die Endstufe wird durch die Transistoren T_3 und T_4 gebildet. Wenn T_2 gesperrt ist (Kollektorpotential an T_2 hoch), wird T_3 leitend und T_4 sperrt. Am Ausgang y erscheint hohes Potential („L“). Ist T_2 leitend, so sinkt das Basispotential an T_3, er sperrt. Das an T_4 ansteigende Basispotential führt T_4 in den leitenden Zustand. An y erscheint niedriges Potential („Ø“).

Die in MOS-Technologie ausgeführten logischen Grundschaltungen enthalten als Elemente nur Feldeffekttransistoren. Wird die Möglichkeit, auch Komplementärtransistoren zu integrieren, genutzt (CMOS), so kann man einen Negator aufbauen, bei dem im statischen Zustand nur Sperr-

ströme fließen (Abb. 8.2.4–4). Liegt x auf niedrigem Potential (≈ 0 Volt), sperrt T_1 und T_2 leitet. Der Ausgang y liegt dann auf hohem Potential. Nimmt x dagegen hohes Potential an ($\approx U_B$), sperrt T_2 und T_1 leitet. Am Ausgang y entsteht niedriges Potential. Da stets einer der beiden Transistoren sperrt, können höchstens beim Umschalten beide Transistoren momentan gleichzeitig leitend werden. Es läßt sich zeigen, daß sich jede beliebige Verknüpfung zweier binärer Signale mit Hilfe einer einzigen logischen Grundfunktion darstellen läßt. Da für TTL-Bausteine die NAND-Funktion die Grundschaltung ist (vgl. Abb. 8.2.4–3), sind in Abb. 8.2.4–5 die in Abb. 8.2.4–2 eingeführten logischen Funktionen durch Kombinationen der NAND-Grundschaltungen dargestellt.

8.2.4.3. Kombinatorische Schaltungen. Schaltungen, in denen binäre digitale Signale durch logische Funktionen verknüpft werden, bestehen aus einer Reihe von Grundschaltungen, die sich häufig wiederholen und die zu komplizierteren Systemen zusammengesetzt werden können. Die Arten der Zusammenschaltungen logischer Grundfunktionen können in zwei Gruppen eingeteilt werden. Zur ersten Gruppe zählen die *kombinatorischen Schaltungen*. Sie zeichnen sich dadurch aus, daß die Ausgangsgrößen vom augenblicklichen Zustand der Eingangsgrößen abhängen. Typisch für sie ist die Existenz von nur einer Signalflußrichtung. Es gibt keine Rückführungen von Ausgängen auf davorliegende Eingänge logischer Grundschaltungen. Vertreter dieser Schaltungsgruppe sind die als Halbaddierer und Volladdierer bezeichneten Addierstufen. Auch Schaltungen, die Dualzahlen in Dezimalzahlen umwandeln und deren Anzeige in Sieben-Segment-Elementen bewirken, gehören zu den kombinatorischen Schaltungen. Die zweite Gruppe sind die *sequentiellen* oder *Folgeschaltungen*. Hier sind die Ausgangsgrößen von den augenblicklichen und den früher vorhanden gewesenen Eingangsgrößen abhängig. Sequentielle Schaltungen sind also speicherhaltige Schaltungen. Sie zeichnen sich durch zwei Signalflußrichtungen aus. Es gibt Rückführungen von Ausgängen auf weiter vorn liegende Eingänge. Grundelemente sequentieller Schaltungen bilden die bistabilen Kippstufen oder Flipflops.

Addierer für zwei Dualziffern. Bei der Addition zweier einstelliger Dualzahlen x_1 und x_2 können vier Fälle eintreten. Die Summe $\sum$ kann einstellig oder zweistellig sein. Im letzten Fall ist ein Übertrag Ü an die nächsthöhere Stelle zu übergeben. Eine Anordnung, die diese Funktion erfüllt, heißt *Halbaddierer*. Ihre Wertetabelle ist in Tab. 8.2.4–6 angegeben.

Vergleicht man die Wertetabelle mit denen der logischen Grundfunktionen aus Tab. 8.2.4–2, so entnimmt man, daß der Übertrag Ü aus der UND-Verknüpfung der Eingangs-

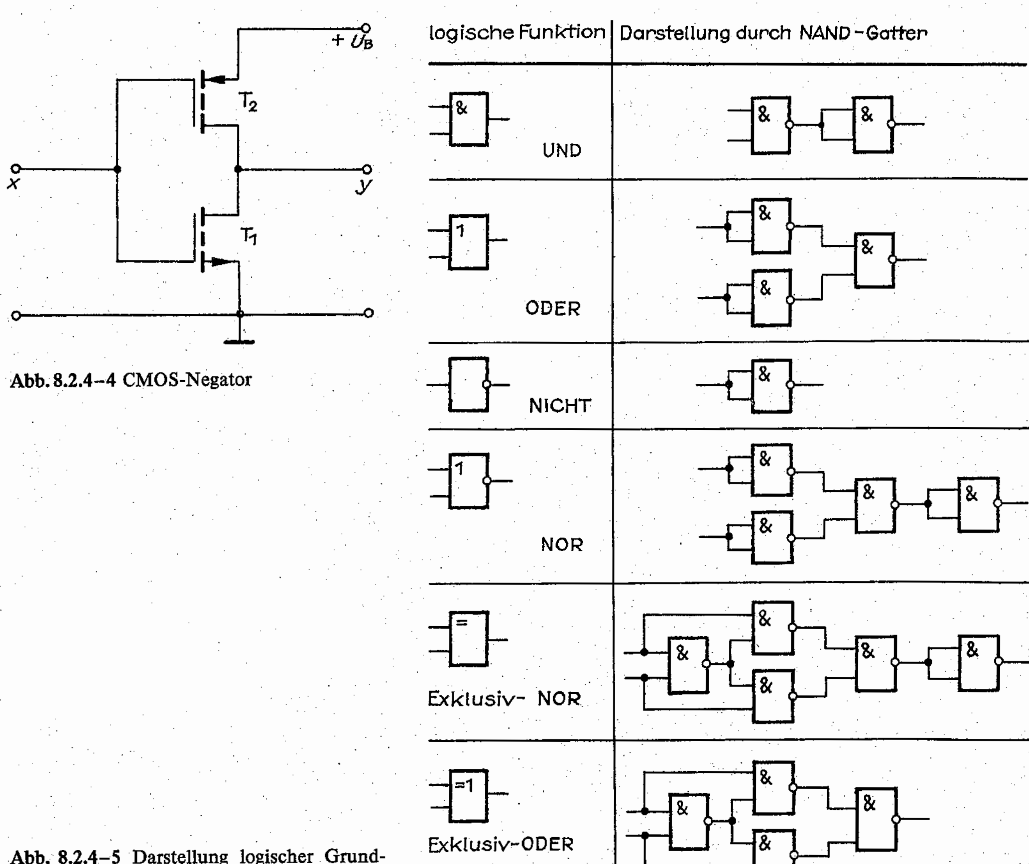


Abb. 8.2.4–4 CMOS-Negator

Abb. 8.2.4–5 Darstellung logischer Grundfunktionen durch NAND-Gatter

größen x_1 und x_2 zustande kommt, während die Summe $\sum$ durch das Exklusiv-ODER gebildet wird. Damit ergibt sich für den Halbaddierer die Schaltung in Abb. 8.2.4–7a. Die Umwandlung der Schaltung, so daß nur noch NAND-Gatter enthalten sind, erfolgt mit Hilfe von Tab. 8.2.4–5 und ergibt Abb. 8.2.4–7b. Ein Halbaddierer kann bei der Addition mehrstelliger Dualzahlen nur die Stelle mit dem niedrigsten Stellenwert einnehmen, da er keinen Übertrag der vorangegangenen Stelle aufnimmt. Diese Aufgabe erfüllt der *Volladdierer*. Er hat drei Eingänge, nimmt also zu den beiden zu addierenden Dualziffern noch den Übertrag der nächstniedrigeren Stelle auf und bildet an seinen beiden Ausgängen die Summe der drei Eingangsgrößen und den Übertrag für die Stelle mit dem nächsthöheren Stellenwert. Mit diesen Eigenschaften kann der Volladdierer für Ziffern mit beliebigem Stellenwert eingesetzt werden. Aus seiner Wertetabelle folgt die Zusammensetzbarkeit aus zwei Halbaddierern und einer Disjunktion (Abb. 8.2.4–8). Aus einem Halbaddierer und drei Volladdierern läßt sich eine Additionsschaltung für vierstellige Dualzahlen aufbauen. Diese eignen sich gut zur Verschlüsselung von Dezimalzahlen. In diesem Zusammenhang bezeichnet man vierstellige Dualzahlen als *Tetraden*. Jeder Dezimalziffer kann nämlich eine Tetrade zugeordnet werden. Diese Darstellung hat gegenüber der Darstellung im Stellenwertsystem der Dualzahlen den Vorteil größerer Übersichtlichkeit. Der Dezimalzahl 169 entsprechen dann die drei Tetraden ØØØL ØLLØ LØØL.

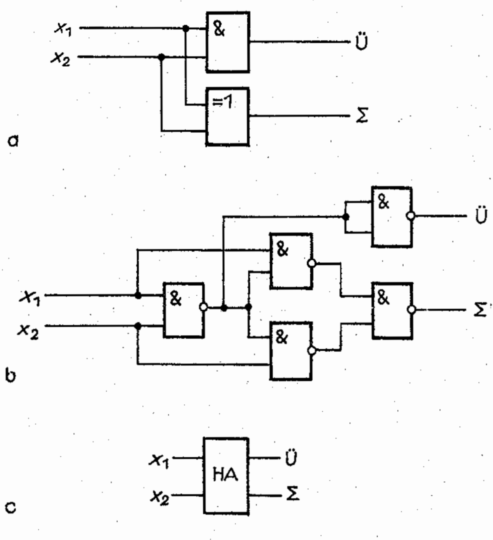


Abb. 8.2.4–7 Halbaddierer, (a) mit UND und Exklusiv-ODER, (b) mit NAND, (c) Symbol

Tab. 8.2.4–6 Wertetabelle eines Halbaddierers

x_1	x_2	$\sum$	Ü
Ø	Ø	Ø	Ø
Ø	L	L	Ø
L	Ø	L	Ø
L	L	Ø	L

x_1	x_2	x_3	Σ	Ü
0	0	0	0	0
0	0	L	L	0
0	L	0	L	0
0	L	L	0	L
L	0	0	L	0
L	0	L	0	L
L	L	0	0	L
L	L	L	L	L

a

Abb. 8.2.4–8 Volladdierer, (a) Wertetabelle, (b) Aufbau, (c) Symbol

Abb. 8.2.4–9 Addition zweier vierstelliger Dualzahlen

Da mit vier Dualstellen sechzehn Zahlen gebildet werden können, bleiben sechs Kombinationen ungenutzt. Diese Art der Zuordnung von Dualzahlen zu Dezimalzahlen nennt man BCD-Kode (*b*inär *c*odierte *D*ezimalzahl). In Abb. 8.2.4–9 ist eine Additionsschaltung für vierstellige Dualzahlen angegeben.

BCD-7-Segment-Decodierer. Beispiele für kombinatorische Schaltungen sind auch Codier- und Decodiernetzwerke. So werden z. B. in Taschenrechnern und Digitaluhren Dezimalziffern durch eine 7-Segment-Anzeige (Abb. 8.2.4–10) mit Hilfe von Flüssigkristallen (vgl. 8.1.6.9.) dargestellt. Bezeichnet man jedes Segment mit einem Buchstaben, so lassen sich in einer Tetrade binär codierte Dezimalziffern, so der Ansteuerung der entsprechenden Segmente zuordnen, daß durch die sichtbar gemachten Segmente die entsprechende Dezimalziffer abgelesen werden kann. Die zugehörige Wertetabelle (Tab. 8.2.4–11) enthält vier Eingänge für die binär codierten Dezimalzahlen und sieben Ausgänge für die entsprechenden Segmente der Ziffernanzeige. Auf der Ausgangsseite soll „L" die Sichtbarmachung eines Segments bedeuten. Solche BCD-7-Segment-Decodierer sind als integrierte Schaltkreise erhältlich.

8.2.4.4. Sequentielle Schaltungen. In sequentiellen oder Folgeschaltungen ist der Zustand der Ausgangsgrößen nicht nur vom augenblicklichen Zustand der Eingangsgrößen, sondern auch von den früher vorhanden gewesenen Eingangsgrößen abhängig. Sequentielle Schaltungen enthalten Speicher, die früher vorhandene Eingangssignale mit zeitlich darauffolgenden verknüpfen. Rückführungen von Ausgängen auf davorliegende Eingänge führen zu zwei Signalflußrichtungen. Einfachste Formen von sequentiellen Schaltungen sind Flipflops.

RS-Flipflops. Neben den elementaren logischen Gattern (vgl. 8.2.4.1.) gehören Flipflops zu den wichtigsten Grundelementen digitaler Schaltungen. Sie haben mindestens zwei Eingänge und zwei Ausgänge. Flipflops können zwei stabile Zustände einnehmen, bei denen die Ausgänge zueinander invers belegt sind. Das einfachste Flipflop ist das RS-Flipflop, das aus zwei NAND-funktionen (vgl. 8.2.4.2.) besteht, die über Rückführungen miteinander verbunden sind (Abb. 8.2.4–12). Die bemerkenswerte Eigenschaft der Flipflops ist, daß sie ein „Gedächtnis" besitzen, sich also speicherartig verhalten. Die Zustandstabelle zeigt, welche Ausgangsbelegungen sich ergeben, wenn die Eingänge R und S der Reihe nach wie angegeben belegt werden. Alle Zustände sind mit der NAND-Zustandstabelle in 8.2.4.1. verträglich. Die Herbeiführung des Zustands $Q = \mathrm{L}$, $\bar{Q} = \emptyset$ durch $S = \mathrm{L}$, $R = \emptyset$ wird als „Setzen" bezeichnet. Beim Entfernen dieser Eingangsbelegung und Anlegen von $S = \mathrm{L}$, $R = \mathrm{L}$ bleibt dieser Zustand erhalten. Es sei daran erinnert, daß dieser Zustand beim jeweiligen NAND-Glied einem unbelegten Emitter des Mehremittertransistors in Abb. 8.2.4–3 entspricht und damit keine elektrische Veränderung bewirkt. Dieser Zustand ist daher mit „Speichern" zu bezeichnen. Entsprechendes gilt für den durch die Eingangsbelegung $R = \mathrm{L}$, $S = \emptyset$ herbeiführbaren Zustand „Rücksetzen". Die Eingangsbelegung $R = \emptyset$, $S = \emptyset$ führt zu einer nichtinversen Belegung der beiden Ausgänge und lie-

Tab. 8.2.4–11 Wertetabelle des BCD-7-Segment Decodierers

	x_4	x_3	x_2	x_1	A	B	C	D	E	F	G
0	Ø	Ø	Ø	Ø	L	L	L	L	L	L	Ø
1	Ø	Ø	Ø	L	Ø	L	L	Ø	Ø	Ø	Ø
2	Ø	Ø	L	Ø	L	L	Ø	L	L	Ø	L
3	Ø	Ø	L	L	L	L	L	L	Ø	Ø	L
4	Ø	L	Ø	Ø	Ø	L	L	Ø	Ø	L	L
5	Ø	L	L	L	L	Ø	L	L	Ø	L	L
6	Ø	L	Ø	Ø	L	Ø	L	L	L	L	L
7	Ø	L	L	L	L	L	L	Ø	Ø	Ø	Ø
8	L	Ø	Ø	Ø	L	L	L	L	L	L	L
9	L	Ø	L	L	L	L	L	L	Ø	L	L

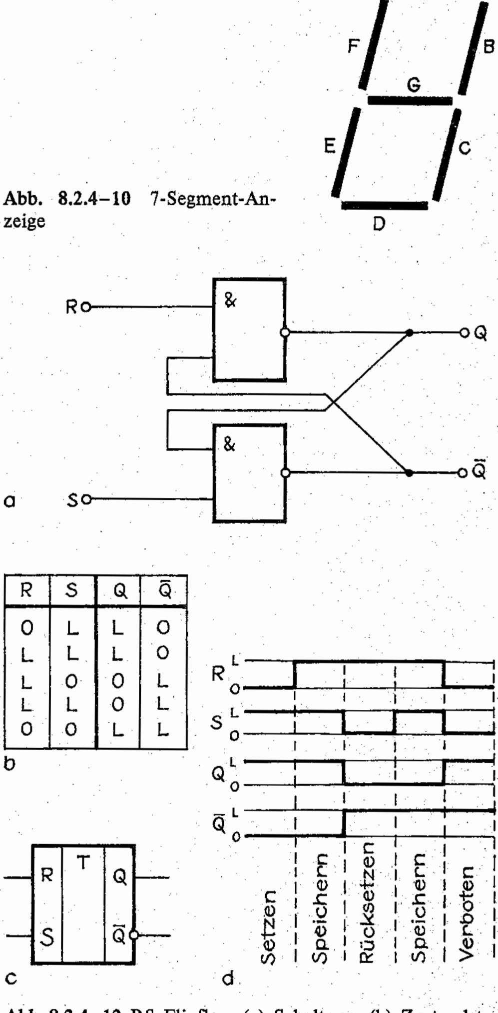
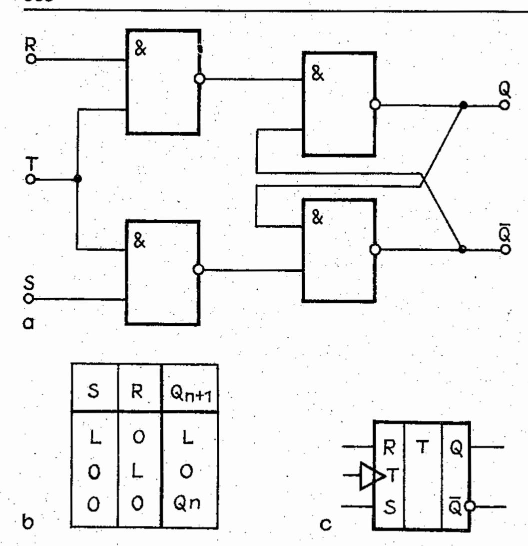


Abb. 8.2.4–12 RS Flipflop, (a) Schaltung, (b) Zustandstabelle, (c) Schaltzeichen, (d) Impulsdiagramm



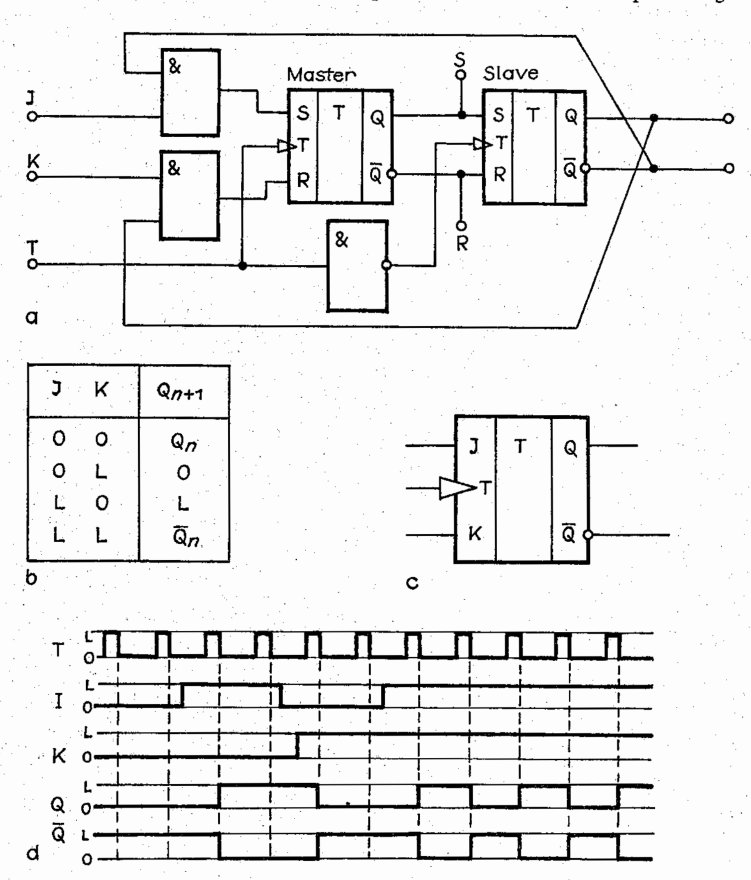
S	R	Q_{n+1}
L	O	L
O	L	O
O	O	Q_n

fert nach Abschaltung der Eingangssignale keinen definierten Speicherzustand. Diese Belegung ist daher nicht zugelassen. Das RS-Flipflop stellt einen Speicher für 1 bit dar und ist das Grundelement für statische Speicher.

Abb. 8.2.4–13 Getaktetes RS-Flipflop, (a) Schaltung, (b) Zustandstabelle, (c) Schaltzeichen

Getaktete RS-Flipflops. In Systemen mit Taktsteuerung sind nur solche Flipflops einsetzbar, die an ihren Eingängen liegende Signale nur zu bestimmten Zeitabschnitten übernehmen, ansonsten aber „geschlossen" sind. Diese Eigenschaft erzeugt man durch die Vorschaltung zweier weiterer Gatter (Abb. 8.2.4–13a). Wenn am Takteingang *T* Ø anliegt, haben die an *R* und *S* anliegenden Signale keine Wirkung auf das Flipflop. Erst wenn sich an *T* hohes Potential („L") befindet, übernimmt das Flipflop die Information an *R* und *S*. Durch die vorgeschalteten Inverter kehrt sich auch die Zuordnung der Eingangsbelegungen zu den entsprechenden Ausgangsbelegungen gegenüber dem statischen RS-Flipflop um. Nach Durchlauf eines Taktimpulses ergibt sich jetzt eine Belegung an Q entsprechend der Zustandstabelle Abb. 8.2.4–13b. Nun ist der Eingangszustand $S = \mathrm{L}$, $R = \mathrm{L}$ auszuschließen.

J-K-Master-Slave-Flipflops. Dieses Flipflop (Abb. 8.2.4–14) besteht aus zwei in Kette geschalteten, getakteten RS-Flipflops, bei denen überkreuzte Rückführungen zwischen den Ausgängen des zweiten und den Eingängen des ersten Flipflops bestehen. Das erste Flipflop heißt Master (Herr), das zweite Slave (Knecht, Sklave). Zwischen beiden wird der Takt invertiert. Damit erfolgen Aufnahme und die Speicherung eines Signalzustands zeitlich nachein-



J	K	Q_{n+1}
O	O	Q_n
O	L	O
L	O	L
L	L	$\bar{Q}_n$

Abb. 8.2.4–14 J-K-Master-Slave-Flipflop, (a) Schaltung, (b) Zustandstabelle, (c) Schaltzeichen, (d) Impulsdiagramm

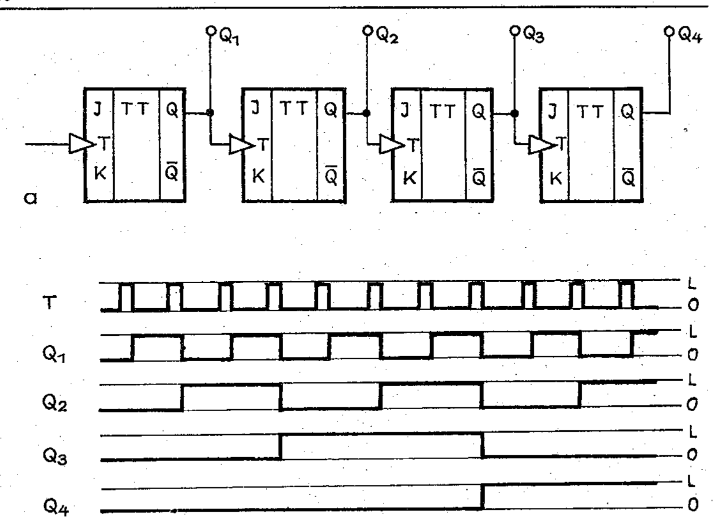


Abb. 8.2.4–15 Asynchronzähler, (a) Schaltung, (b) Impulsdiagramm

ander. Bei ansteigender Flanke des Taktimpulses öffnet der „Master" und stellt an seinen Ausgängen den Zustand her, der seiner Eingangsbelegung nach der Zustandstabelle Abb. 8.2.4–13b zugeordnet ist. Mit abfallender Taktflanke schließt der „Master" und verarbeitet keine neue Information. Durch die Inversion des Taktzustandes öffnet sich der „Slave", übernimmt die am Master-Ausgang liegende Belegung und stellt die entsprechende Ausgangsbelegung her. Durch die beiden Rückführungen wurde die bisher verbotene Eingangsbelegung beseitigt. Die Zustandstabelle gibt für jede mögliche Belegungskombination der Signaleingänge J und K ein definiertes Verhalten des Flipflops. Liegt an den Signaleingängen Ø, Ø, so bleibt nach Durchlauf eines Taktimpulses der vorherige Zustand bestehen. Liegt an $J = Ø$ und $K = L$, so ist nach Durchlauf eines Taktimpulses, unabhängig davon, welchen Zustand der Ausgang vor dem Taktimpuls hatte, am Q-Ausgang Ø. Entsprechendes gilt für die dazu inverse Belegung der Signaleingänge. Liegen die beiden Signaleingänge auf hohem Potential („L"), so wird nach Durchlauf eines Taktimpulses die vorher vorhanden gewesene Ausgangsbelegung invertiert. Abb. 8.2.4–14b stellt das Impulsdiagramm des J-K-Master-Slave-Flipflops dar. Die beiden mittleren Zeilen der Zustandstabelle werden im Schieberegister angewendet, die letzte Zeile bei asynchronen Impulszählern.

Zähler. Mit J-K-MS-Flipflops als Grundelemente lassen sich Impulszähler aufbauen. Abb. 8.2.4–15 zeigt einen vierstufigen, asynchronen Dualzähler. Die zu zählenden Impulse werden dem Takteingang des Flipflops mit dem niedrigsten Stellenwert zugeführt. Der Takteingang des Flipflops mit dem nächsthöheren Stellenwert wird vom Ausgang Q des vorangegangenen gespeist. Die Signaleingänge der Flipflops sind unbelegt. Die elektrische Wirkung ist dieselbe, wie beim Anlegen hohen Potentials („L") und entspricht damit der letzten Zeile der Zustandstabelle des J-K-Master-Slave-Flipflops (vgl. Abb. 8.2.4–14b). Nach jedem Taktimpuls wird die Ausgangsbelegung gegenüber der vorangegangenen invertiert.

Werden die Zählimpulse allen Flipflops gleichzeitig zugeführt, spricht man von *Synchronzählern*. Damit hierbei nicht alle Stufen gleichzeitig umschalten, sind die J- und K-Eingänge in geeigneter Weise zu beschalten, so daß in Situationen, in denen das Flipflop nicht umschalten soll, an den Signaleingängen Ø,Ø liegt (erste Zeile der Zustandstabelle Abb. 8.2.4–14b). Mit Hilfe von BCD-7-Segment-Decodierern lassen sich die Zählergebnisse in Ziffern darstellen.

Frequenzteiler. Die oft zu lösende Aufgabe, Impulsfrequenzen zu untersetzen, kann für Teilverhältnisse $2^n:1$ ebenfalls mit Anordnungen der Abb. 8.2.4–15 bewältigt werden. So sind, wie aus dem Impulsdiagramm hervorgeht, beispielsweise nach acht eingegebenen Taktimpulsen am Ausgang Q genau zwei Impulse gezählt worden. Die Frequenzuntersetzung in Teilverhältnisse, die ungeradzahlige Faktoren enthalten, läßt sich durch geeignete Verbindungen von Ausgängen mit Signaleingängen herstellen. Maßnahmen dieser Art müssen in quarzgesteuerten elektronischen Uhren vorgenommen werden (vgl. 8.1.5.2.). Abb. 8.2.4–16 zeigt einen Frequenzteiler 3:1.

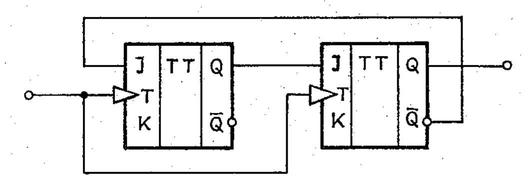


Abb. 8.2.4–16 Frequenzteiler 3:1

Schieberegister sind Speicher, in denen die bereits enthaltene Information einen Speicherplatz weiter rückt, wenn eine neue Information aufgenommen wird. Sie lassen sich ebenfalls aus Master-Slave-Flipflops aufbauen. Zum unveränderten Weiterschieben eines Zustands können die beiden mittleren Zeilen der Zustandstabelle Abb. 8.2.4–14b dienen, bei denen die Signaleingänge J und K eine jeweils zueinander inverse Belegung aufweisen. Abb. 8.2.4–17 zeigt die Schaltung eines Schieberegisters und das dazugehörige Impulsdiagramm, wenn die Information „L" durchgeschoben wird. Die jeweils invertierte Belegung von J und K wird durch Zwischenschaltung eines Negators erreicht.

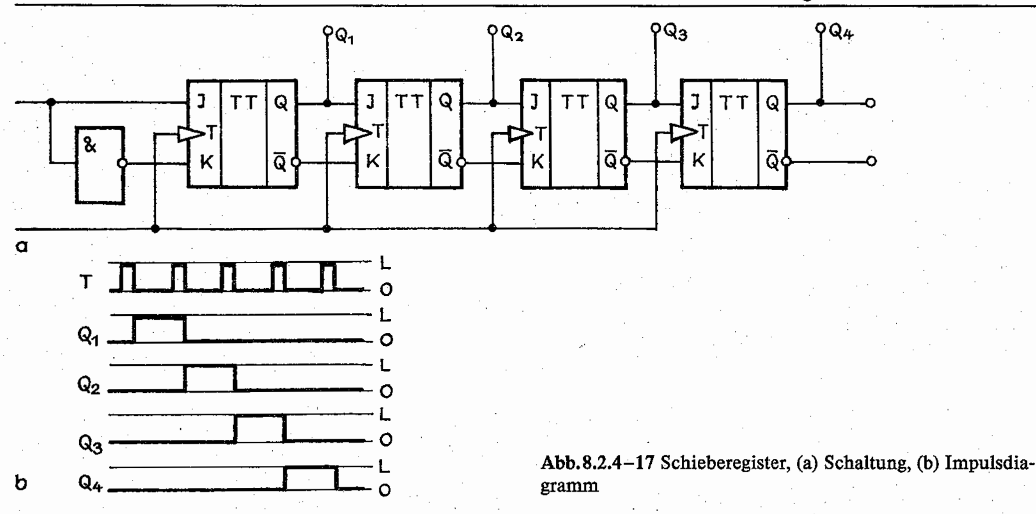


Abb. 8.2.4-17 Schieberegister, (a) Schaltung, (b) Impulsdiagramm

8.2.5. Analog-Digital-Umsetzer

Der fortgeschrittenen Anwendung der digitalen Signalverarbeitung steht der Umstand entgegen, daß viele Meßwerte und Signale kontinuierliche Größen sind. Die Verbindung zwischen analog vorliegenden Signalen und der digitalen Verarbeitung schaffen die Analog-Digital-Umsetzer (ADU), sie verwandeln ein analoges Eingangssignal, z.B. eine Spannung U, wie im Digitalvoltmeter, in eine dazu proportionale Zahl, das digitale Ausgangssignal. Diese Umwandlung wiederholt sich in bestimmten Zeitintervallen. Dazu ist die zeitliche Abtastung des analogen Eingangssignals, die Quantisierung der Signalamplitude und die Verschlüsselung des ermittelten Amplitudenwertes erforderlich.

Zu den Grundprinzipien der Analog-Digital-Wandlung gibt es vielfältige Schaltungsvarianten. In Abb. 8.2.5-1 ist ein einfacher, nach dem Zählverfahren arbeitender ADU angegeben, der im Gegensatz zu anderen Prinziplösungen nur eine Referenzspannung benötigt, aber n Rechenschritte braucht, wenn n die größte darstellbare Zahl ist. Daher ist die Umwandlungszeit, die zwischen 1 und 100 ms liegt, relativ groß. Die wichtigsten Bestandteile sind ein Kippspannungsgenerator mit zeitlinearem Anstieg der Spannung und Bezugsspannungsquelle, ein quarzgesteuerter Impulsgenerator, Zähl- und Anzeigeanordnungen sowie logische Gatter. Der Kippgenerator erzeugt eine Sägezahnspannung $U_{Kipp}(t) = (U_{Bez}/\tau)t - U_0$, die an beide Komparatoren gelangt. Als Komparatoren, die dem Vergleich zweier Spannungen dienen, arbeiten Operationsverstärker, die bis an ihre Grenzwerte der Ausgangsspannung ausgesteuert sind. Die zu vergleichenden Spannungen werden an den Differenzspannungseingang gelegt. Ist die zu vergleichende Spannung größer als die Bezugsspannung, liegt am Ausgang die Spannung $U_{a\,max}$ an, die der positiven Aussteuerungsgrenze entspricht. Ist die zu vergleichende Spannung dagegen geringer als die Bezugsspannung, nimmt der Ausgang die negative Aussteuerungsgrenze $U_{a\,min}$ an. Der oberen Aussteuerungsgrenze ist der logische Wert L, der unteren der Wert 0 zugeordnet. Solange die Sägezahnspannung negativ ist, sind an beiden unteren Komparatoreingängen

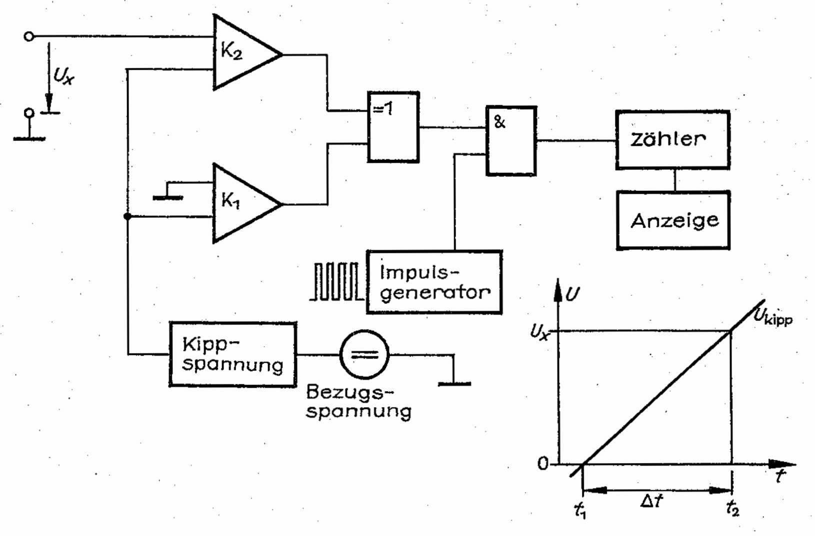


Abb. 8.2.5-1 Analog-Digital-Wandler nach dem Zählverfahren

die Spannungen kleiner als an den oberen. An beiden Komparatorausgängen liegt logisch 0. Da aber die logische Funktion Exklusiv-ODER nur bei ungleicher Eingangsbelegung am Ausgang L zeigt, liegt ein Eingang des darauffolgenden UND-Gliedes auf 0 und dem Zähler wird nichts angeboten. Gelangt die Sägezahnspannung auf positive Werte, nimmt der Komparator die positive Aussteuerungsgrenze, also den logischen Wert L an. Dadurch liegt am Ausgang des Exklusiv-ODER-Gatters logisch L und die Zählimpulse gelangen auf den Zähler und zur Anzeige. Übersteigt die Kippspannung den Wert der zu wandelnden Spannung U_x, schaltet der Komparator K_2 auf L und unterbricht die Zählung. Damit ist der Spannung U_x die in der Anzeige stehende Zahl zugeordnet worden. Wenn vor Zählbeginn die Anzeige auf 0 gestellt wurde, zeigt sie nach Zählabschluß $N = \Delta t \cdot \nu$, wobei Δt das Zeitintervall zwischen Nulldurchgang der Sägezahnspannung und ihrem Erreichen des Wertes der zu wandelnden Spannung U_x ist, und ν die Frequenz des Impulsgenerators darstellt. Nach $(U_{Bez}/\tau)t - U_0$ ist $U_x = (U_{Bez}/\tau)(t_2 - t_1) = (U_{Bez}/\tau)\Delta t$ und $N = \nu \cdot \tau U_x/U_{Bez}$.

Im allgemeinen wird das alte Zahlenergebnis solange in der Anzeige gespeichert, bis das neue zur Verfügung steht.

8.2.6. Hochintegrierte Speicher- und Mikrorechnerschaltkreise

Der sich mit der technologischen Entwicklung erhöhende Integrationsgrad von Halbleiterschaltkreisen erfordert aus wirtschaftlichen Gründen zur Rechtfertigung der hohen Entwicklungskosten auch sehr hohe Stückzahlen. Hochintegrierte Digitalschaltungen, auf die das zutrifft, sind Halbleiterspeicher und Mikrorechnerschaltkreise. Speicher für digitale Signale stellen wesentliche Baugruppen für Systeme der Informationsverarbeitung dar. Sie dienen zur Speicherung von Programmen und Daten in Digitalrechnern und bei der Prozeßsteuerung, sie können zur Umcodierung, als Zeichengeneratoren und vielem anderen verwendet werden. Das große technische Interesse an Speicheranordnungen hat dazu geführt, daß auch eine Anzahl anderer physikalischer Phänomene auf ihre Eignung zur Informationsspeicherung untersucht und auch teilweise schon angewandt werden. Einige Beispiele sind in 8.2.6.1. skizziert. Besondere Bedeutung haben die Halbleiterspeicher gewonnen. Nach ihrem Arbeitsregime kann man die Halbleiterspeicher in Matrixspeicher und Umlaufspeicher untergliedern. In Matrixspeichern sind die Speicherzellen, die im einfachsten Fall je 1 bit, d.h. entweder „L" oder „Ø" speichern können, in einer rechteckigen Matrix angeordnet. Jede Speicherzelle ist eindeutig durch ihre Adresse, d. h. durch Zeilen- und Spaltenzahl gekennzeichnet. Wenn mit jeder Adresse nur eine Speicherzelle erfaßt wird, nennt man den Speicher „bitorganisiert". Sind dagegen mehrere Speicherzellen zu einem „Wort" (etwa mit einer Länge von 4 oder 8 bit) zusammengefaßt, das unter Einsparung von Zuleitungen mit einer Adresse angewählt werden kann, so spricht man von einem „wortorganisierten" Speicher. Innerhalb der Matrixspeicher können weitere Unterscheidungen getroffen werden. Speicher, in die häufig Informationen eingelesen und ausgelesen, gelöscht und neu eingelesen werden, heißen Schreib-Lese-Speicher. Wegen des wahlfreien Zugriffs werden sie als RAM (engl. *R*andom-*A*ccess *M*emory) bezeichnet. Die im RAM enthaltene Information geht bei Ausfall der Versorgungsspannung verloren. Halbleiterspeicher, in die einmalig, im Verlaufe des Herstellungsprozesses eine Information eingeschrieben wurde, die beliebig oft auslesbar ist, heißen Festwertspeicher, Nur-Lese-Speicher oder ROM (engl. *R*ead-*O*nly *M*emory). Ihre Information bleibt bei Ausfall der Versorgungsspannung erhalten.

Die zweite Gruppe der Halbleiterspeicher sind die Umlaufspeicher. Ihnen liegt das Prinzip des Schieberegisters (vgl. 8.2.4.4.) zugrunde, in denen die Information umläuft. Umlaufspeicher bestehen aus parallel zueinander angeordneten Schieberegistern. Ihre Zugriffszeit ist größer als die von Matrixspeichern, da die Information nur am Registerausgang verfügbar ist.

Zu universell einsetzbaren und damit mit hohen Stückzahlen benötigten Schaltkreisen zählen Schaltkreise, die in analoger Form anfallende Meßdaten digitalisieren (AD-Wandler, vgl. 8.2.5.) bzw. digitale Größen in analoge zurückführen (DA-Wandler). Aufgaben dieser Art treten insbesondere bei der Prozeßsteuerung häufig auf.

Als hochintegrierter Schaltkreis universeller Einsetzbarkeit erweist sich der Mikroprozessor, der Rechen- und Steuerwerk eines Mikrorechners enthält (vgl. 8.2.6.4.).

8.2.6.1. Physikalische Prinzipien der Informationsspeicherung. Der Begriff Informationsspeicherung läßt sich sehr weit fassen. Hier müssen Bücher, Schallplatten, Fotos, bespielte Magnettonbänder einbezogen werden; das im Gehirn lokalisierte menschliche Gedächtnis; die sich in der Doppelhelix der DNS reproduzierenden Erbanlagen, der von der Gesellschaft repräsentierte Kenntnisstand im natur- und geisteswissenschaftlichen und im technologischen Bereich müssen ebenfalls der Informationsspeicherung zugeordnet werden. Im engeren Sinne versteht man unter einem Informationsspeicher ein physikalisches System, das Daten, Signale oder Werte physikalischer Größen aufnehmen kann und für einen späteren Zeitpunkt zum Abruf bereit hält. Speicher lassen sich nach verschiedenen Gesichtspunkten ordnen: nach der zur Einschreibung der Information in den Speicher genutzten Energieform (magnetisch, elektrisch, mechanisch, optisch, chemisch), nach der Signalart (analog, digital), nach dem Speichermedium (Ferritkern, dünne magnetische Schicht, fotografische Emulsion, strukturiertes Halbleitermaterial), nach der Art der Beständigkeit der Information (selbsterhaltend oder nur durch Energiezufuhr erhaltbar), oder danach, ob eine ein- oder mehrmalige Speicherung möglich ist (Irreversibilität bzw. Reversibilität des Speichermediums). Neben den in der elektronischen Datenverarbeitung in hohem Maße benutzten bipolaren oder MIS-Halbleiterspeichern bietet die Nutzung einer Reihe physikalischer Effekte Möglichkeiten zur Informationsspeicherung. Wie auch bei den Halbleiterspeichern sind insbesondere digitale Speicher gefragt, deren Speicherelemente jeweils einen von zwei möglichen Zuständen einnehmen, d.h. die binär arbeiten. Zu ihnen gehören die *Kryospeicher*, deren Speicherzustände Supraleitung und Normalleitung sind (vgl. 8.1.4.). Seit der Beschreibung holografischer Abbildungen (vgl. 4.11.1.5.) durch GABOR und der Existenz einer monochromatischen und kohärenten Strahlungsquelle hoher Energiedichte in Form des Lasers (vgl. 4.11.1.5.) sind holografische Speicher dreidimensionaler Objekte auf zweidimensionalen Fotoplatten realisierbar, die beim Wiedergabevorgang dreidimensionale Abbilder erzeugen. Auf synthetischem Wege, z. B. mit Hilfe von Rechenautomaten, können binäre Hologramme hergestellt werden, die sich wieder abtasten und rechnerisch verarbeiten lassen. Als reversible Speichermedien sind lichtemp-

findliche Gläser, Thermoplaste, Ferroelektrika und magnetische Schichten denkbar. Als aussichtsreich wird eine ≈70 nm dicke ferromagnetische Mangan-Bismut-Schicht angesehen. *Hologrammspeicher* ermöglichen die derzeit höchste Speicherdichte, setzen der Anwendbarkeit in der Praxis aber noch beträchtliche Schwierigkeiten entgegen.

Länger bekannt und auch angewandt werden ferromagnetische Speichermedien. Das älteste magnetische Verfahren zur Schallaufzeichnung nach POULSEN verwendete Stahldraht, der am Luftspalt eines mit dem Signalstrom gespeisten ringförmigen Elektromagneten (Aufzeichnungskopf) vorbeigezogen wurde. Der dadurch erzeugte Magnetisierungszustand ist über einen Wiedergabekopf beliebig oft reproduzierbar. Der Ersatz des Stahldrahtes durch ein magnetisierbares Band brachte wesentliche Qualitätsverbesserungen. Gebräuchliche *Magnetbänder* bestehen meist aus einer Polyester-Trägerfolie, auf der sich, in ein Bindemittel eingebettet, die magnetisierbaren Teilchen (Fe_2O_3 und Fe_3O_4 bzw. CrO_2) befinden. Die nadelförmigen Partikeln haben Längen um 1µm. Standardisierte Bänder werden für Spulentonbandgeräte, Kassettenrecorder, Video- und auch in großem Umfang zur Speicherung digitaler Informationen in Magnetbandspeichergeräten eingesetzt. Wenn man die Speicher in Datenverarbeitungssystemen charakterisiert, sind Speicherkapazität und Zugriffszeit wichtige Parameter. Das Magnetband stellt den externen Massenspeicher mit der größten Speicherkapazität, allerdings auch der größten Zugriffszeit dar. Die Zugriffszeit kann bis zu einigen Sekunden betragen.

Eine technische Abwandlung des Prinzips ist der *Magnetplattenspeicher*, eine beiderseits mit einer magnetisierbaren Schicht versehene, schnell rotierende Platte, über der im Abstand von wenigen Mikrometern Magnetköpfe Informationen ein- und auslesen können. Als Bestandteile des Hauptspeichers in Rechenanlagen werden *Ferritkerne* angewandt. Ihre Bedeutung verringert sich allerdings angesichts der Entwicklung hochintegrierter Halbleiterspeicher. Sie haben Ringform, bestehen aus ferrimagnetischen Oxiden mit Spinellstruktur und können mit Hilfe des Magnetfeldes stromdurchflossener Drähte, die durch das Ringinnere geführt werden, entsprechend der Stromrichtung zwischen den beiden möglichen Magnetisierungsrichtungen hin- und hergeschaltet werden. Ferritkerne haben eine angenähert rechteckige Hysteresekurve und behalten eine beträchtliche Remanenz. Sie haben den Vorteil aller magnetischen Speicher, den der Nichtflüchtigkeit der Information bei Ausfall der Stromversorgung. Ferritkerne werden zu quadratischen Speichermatrizen zusammengeführt, wobei durch jeden Ferritkern entsprechend der gewählten Ansteuerung vier oder sechs Drähte zu führen sind. Eine Matrix mit 128 × 128 Ferritkernen von 1,3 mm Durchmesser hat ein Ausmaß von ≈ 18 × 18 cm². Ferritkernspeicher haben Zugriffszeiten zwischen 1 und 100 s.

Anregung für die Suche nach neuartigen Speicherprinzipien war das Bedürfnis nach größeren Speicherkapazitäten auf immer geringeren Flächen und großer Zuverlässigkeit. Als Festkörperspeicher großer Kapazität auf der Grundlage des Schieberegisterprinzips ist der *Magnetblasenspeicher* (Bubble-Speicher) eine auch in der Praxis interessante Variante. In einem Schieberegister (vgl. 8.2.4.4.) wird die am Eingang eingegebene Information bei jedem Takt um einen Speicherplatz weitergeschoben. Am Ausgang des Registers besteht Zugriff zu ihr und sie wird gelesen und erneut dem Eingang des Registers zugeführt (Umlaufspeicher). Beim Magnetblasenspeicher können in einer dünnen magnetischen Schicht infolge eines senkrecht zur Schichtebene angelegten magnetischen Feldes geeigneter Größe zylinderförmige magnetische Domänen stabil gehalten werden. Ihre Magnetisierungsrichtung ist der der restlichen Schicht entgegengerichtet. Die zylindrischen magnetischen Domänen

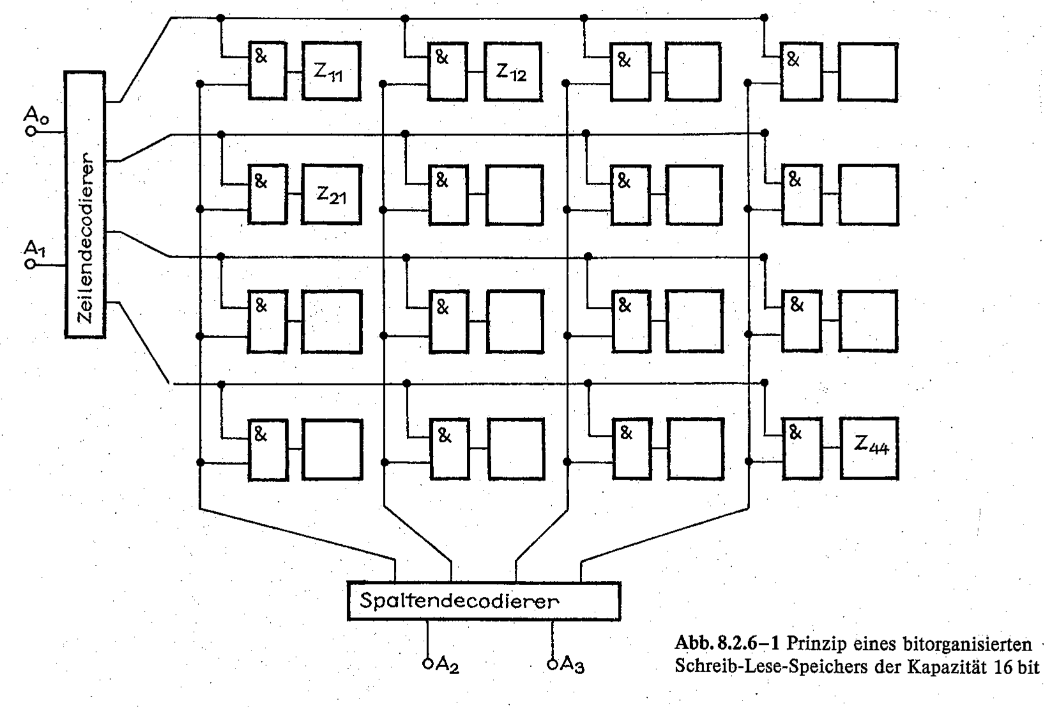


Abb. 8.2.6–1 Prinzip eines bitorganisierten Schreib-Lese-Speichers der Kapazität 16 bit

sind unter der Wirkung von Magnetfeldern in der Schichtebene verschiebbar. Der Anwesenheit bzw. dem Fehlen einer Domäne an einer bestimmten Stelle der Schiebekette ist dann die Information „L" bzw. „Ø" zugeordnet. Es ergeben sich Speicherkapazitäten zwischen 10^6 und 10^8 bit und Zugriffszeiten von wenigen Millisekunden. Für die dünnen magnetischen Schichten werden meist magnetische Granate, das sind ferrimagnetische Oxide mit kubischer Kristallstruktur, verwendet.

In die Reihe der Speicher, die nach dem Prinzip des Schieberegisters arbeiten, sind auch Zusammenschaltungen von ladungsgekoppelten Bauelementen (CCD) einzuordnen (vgl. Abb. 8.1.6-13). Von Magnetblasenspeichern und ladungsgekoppelten Schaltungen werden Speicherkapazitäten bis zu 10^6 bit/Chip erwartet.

8.2.6.2. Schreib-Lese-Speicher (RAM). Wenn bei einem Schreib-Lese-Speicher die Adresse eines Speicherplatzes, d.h. seine Zeilen- und Spaltennummer, vorgegeben ist, lassen sich unter dieser Adresse beliebig oft digitale Informationen einlesen und abfragen. Die Speichermatrix besteht aus 2^n Zeilen und 2^m Spalten. Der Aufruf einer bestimmten Speicherzelle erfolgt dadurch, daß an die zugehörigen Zeilen- und Spaltenleitungen ein logisches Signal (z. B. „L") gelegt wird. Wenn die Speicherzellen auf diese Weise direkt von außen angesteuert würden, so müßte der Speicher $2^n + 2^m$-Anschlüsse haben. Für einen Speicher der Kapazität 1 kbit (1024 Speicherplätze) wären das $2^5 + 2^5 = 64$ Anschlüsse. Durch Verwendung von Decodierschaltungen läßt sich diese Anzahl stark reduzieren. Die Decodierschaltungen ermöglichen die Eingabe der Dualzahl der anzuwählenden Spalten- bzw. Zeilenzahl. Auf diese Weise reduziert sich die Anzahl der nötigen Außenanschlüsse für die Adressenleitungen auf $n + m$, d. h. für einen RAM mit 1024 Speicherplätzen auf 10. Abb. 8.2.6-1 zeigt das Prinzip der Speichermatrix eines 16-bit-RAM. Außer der Speichermatrix und den Decodierern gibt es in einem Halbleiterspeicher noch Schreib-Lese-Verstärker und eine Steuerschaltung, die für die Herstellung der Betriebsart „Lesen" oder „Schreiben" zu sorgen hat. Ist der herausgeführte Anschluß der Steuerlogik, der den Betriebszustand „Schreiben" herbeiführt, mit $\overline{\text{WE}}$ (engl. *w*rite *e*nable, d.h. schreibbereit) bezeichnet, so bedeutet das, daß bei Anlegen von „Ø" an diesen Anschluß über den Dateneingang die gewünschte Information in die angesteuerte Zelle eingeschrieben werden kann. Außer in der Schreibphase befindet sich das RAM im Lesezustand. Nach Ablauf der Zugriffszeit liegt die in der angesteuerten Speicherzelle enthaltene Information am Datenausgang. Schreib-Lese-Speicher, die bei vorhandener Betriebsspannung keiner „Auffrischung" der gespeicherten Information bedürfen (statische RAMs) werden sowohl in bipolarer wie auch in MIS-Technologie hergestellt. Abb. 8.2.6-2 zeigt das Flipflop einer Speicherzelle eines in bipolarer Technik realisierten Schreib-Lese-Speichers. Die für die TTL-Technik typischen Mehremitter-Bipolar-Transistoren kennzeichnet, daß unter den Emittern die mit niedrigem Potential für die elektronische Funktion bestimmend sind.

Im Gegensatz zu statischen Schreib-Lese-Speichern bedürfen dynamische RAMs einer periodischen Auffrischung der gespeicherten Information, da diese als Ladung eines MOS-Kondensators realisiert ist. Dynamische Speicherzellen sind kleiner als statische, ergeben also bei gleichem Integrationsgrad eine höhere Speicherkapazität. Ökonomisches Interesse stimuliert die Anstrengungen zur Weiterentwicklung der technologischen Verfahren. Ziele sind Senkung der Kosten je bit, Reduzierung der Zugriffszeit und Senkung des Leistungsverbrauches. Industriell beherrscht werden bereits (1984) dynamische Schreib-Lese-Speicher mit einer Chipkapazität von 64 kbit mit Zugriffszeiten von Zehntel Mikrosekunden, einem Leistungsverbrauch von 10 µW/bit und Flächen von 60 µm × 60 µm für eine Speicherzelle (Durchmesser eines Haares ≈50 µm). Trotz der Kleinheit dieser Dimension ist die Fläche einer Speicherzelle noch nahezu um zwei Größenordnungen größer als die physikalisch-technische Grenze.

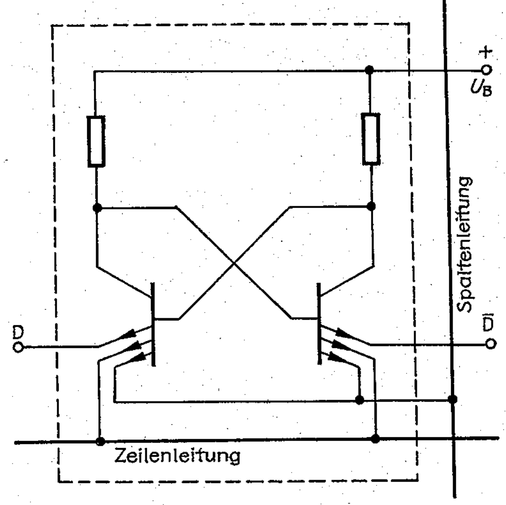


Abb. 8.2.6-2 Speicherzelle eines bipolaren Schreib-Lese-Speichers, D: Datenein- und -ausgang, $\overline{\text{D}}$: zu D invertierter Zustand

8.2.6.3. Festwertspeicher. Müssen unveränderliche Daten jederzeit abrufbereit sein, werden Festwertspeicher eingesetzt. Sie sind fest programmiert, ihr Inhalt kann nur gelesen werden (Nur-Lese-Speicher, Read-Only Memory, ROM). Festwertspeicher dienen u. a. zur Erzeugung von Konstanten und Funktionstabellen, als Zeichengeneratoren, zur Kalenderspeicherung bei Quarzuhren und als Programmspeicher in Mikroprozessorsystemen (vgl. 8.2.6.4.). Festwertspeicher haben einen ähnlichen Aufbau wie Schreib-Lese-Speicher. Die Zellen der Speichermatrix sind jedoch wegen ihres vorgegebenen Inhalts wesentlich einfacher. Die Verknüpfungen zwischen den Zeilen- und Spaltenleitungen, die durch die Programmierung fest eingestellt werden, sind entweder niederohmig, was dem Zustand „Ø" zugeordnet wird, oder hochohmig (Zustand „L"). Das vorgesehene Bitmuster in der Speichermatrix wird bei der Herstellung des Festwertspeichers durch Anwendung einer entsprechenden Maske bewirkt (Maskenprogrammierung). Für den Anwender kann es vorteilhaft sein, einen Festwertspeicher selbst programmieren zu können. Das ist möglich, wenn am fertigen Schaltkreis von außen festgelegt werden kann, für welche Speicherzellen zwischen Zeilen und Spaltenleitungen hochohmige und für welche niederohmige Verbindungen bestehen sollen. In einem Speicherschaltkreis, der nur niederohmige Verbindungen hat, werden für die gewünschten Speicherzellen hochohmige Verbindungen durch das Durchbrennen von als Sicherung wirkenden Ver-

bindungsleitungen erreicht. Meist wird durch einen kurzen Stromstoß eine NiCr-Leitung geschmolzen.

Das Herauslesen der Information aus einer Zelle des Festwertspeichers entspricht dann dem Feststellen des programmierten Zustandes. Bezüglich der Adressencodierung besteht kein prinzipieller Unterschied zum RAM. ROMs sind aber immer wortorganisiert, d. h. durch eine Adresse wird immer eine Bitgruppe, häufig 8 bit, ausgelesen.

Es gibt auch Festwertspeicher, die reversibel programmiert werden können. Sie arbeiten mit MIS-Transistoren, die eine einstellbare Schwellspannung aufweisen. Die Löschung des vorherigen Speicherinhalts erfolgt bei diesen Speichern mit ultraviolettem Licht, mit Röntgenstrahlen oder auf elektrischem Wege. Elektrisch programmierbare und elektrisch löschbare Festwertspeicher werden seit Beginn der 80er Jahre industriell mit einer Speicherkapazität bis zu 64 kbit je Chip hergestellt.

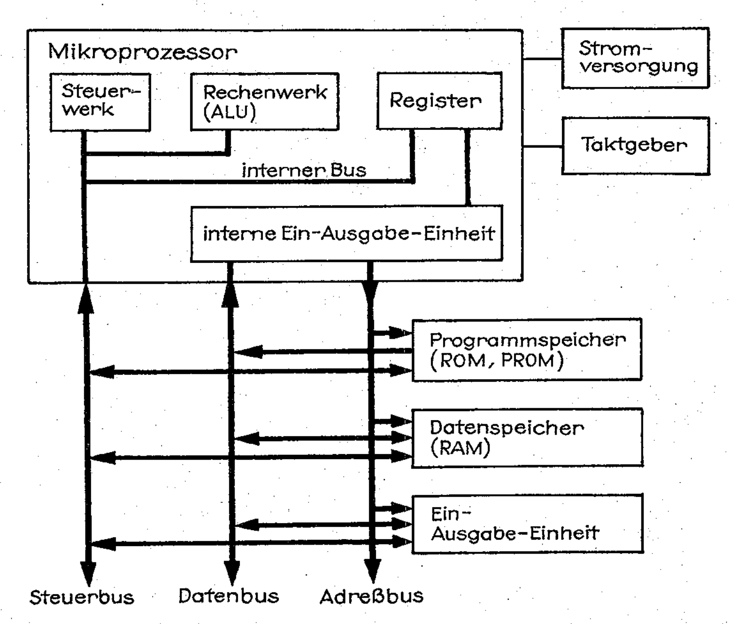


Abb. 8.2.6–3 Prinzipaufbau eines Mikrorechners

8.2.6.4. Mikroprozessoren. Hochintegrierte Schaltkreise mit mehr als 1000 Transistorfunktionen je Chip herzustellen, wird erst dann ökonomisch vertretbar, wenn die Entwicklungs- und Herstellungskosten durch hohe Stückzahlen zu rechtfertigen sind. Allgemein ist jedoch wachsender Integrationsgrad mit zunehmend spezieller werdendem Anwendungszweck gekoppelt. Für spezielle Anwendungen mit hohen Stückzahlen haben sich nur Speicherschaltkreise und Schaltkreise für Uhren und Taschenrechner durchgesetzt. Ein hochintegrierter Schaltkreis, der demgegenüber für universelle Anwendungen einsetzbar ist, muß sich an die jeweilige Aufgabe anpassen lassen. Das ist mit einem Schaltkreis möglich, der aufgrund von außen eingegebener Daten und Befehle eine größere Zahl von Verarbeitungsschritten ausführen kann. Diese Arbeitsweise entspricht der einer zentralen Prozessoreinheit (Verarbeitungseinheit) eines Digitalrechners. Mikroelektronische Schaltkreise mit dieser Eigenschaft werden als Mikroprozessoren bezeichnet. Sie wurden erstmalig 1971 hergestellt. Bereits 1976 befanden sich über 100 Typen in der Fertigung. Ende 1980 wurde die weltweit im Einsatz befindliche Anzahl von Mikroprozessoren auf ≈250 Millionen geschätzt. Um die Jahrtausendwende wird erwartet, daß ihre Anzahl die Milliardengrenze weit überschritten haben wird. Die grundlegende Bedeutung von Mikroprozessoren rechtfertigt, seit ihrer Realisierung von einer neuen Entwicklungsphase der Elektronik zu sprechen, die in ihren Auswirkungen mit der Entdeckung des Transistoreffektes vergleichbar ist. Der Mikroprozessor ist als Teil des Mikrorechners (Abb. 8.2.6–3) aufzufassen. Dabei versteht man unter Mikrorechnern Digitalrechner, die eine feste Wortlänge zwischen 4 und 16 bit haben, welche vom Rechner als Einheit erkannt und verarbeitet wird. Zu den Bestandteilen eines Mikrorechners gehört neben dem Mikroprozessor die *Eingabe-Ausgabe-Einheit*, die die Verbindung des Mikrorechners zur Außenwelt darstellt. Über sie erfolgt die Informationsübergabe vom und zum Rechner. Für die zeitliche Synchronisierung der ablaufenden Verarbeitungsschritte sorgt ein Taktgenerator. Die Taktfrequenzen liegen im Bereich einiger Megahertz. Der *Programmspeicher (Befehlsspeicher)* enthält die vom Rechner der Reihe nach zu befolgenden Arbeitsanweisungen. Programmspeicher haben die Form von Festwertspeichern, die entweder vom Hersteller eingestellt (ROM, vgl. 8.2.6.2.) oder vom Anwender einmalig programmiert werden (PROM). Mit dem Auswechseln des Programmspeichers ist die universelle Anpassung an den entsprechenden Anwendungsfall gegeben. Der *Datenspeicher* ist ein Schreib-Lese-Speicher mit direktem Zugriff zu jedem Speicherplatz (RAM, vgl. 8.2.6.3.), in den man beliebig oft Informationen einschreiben und auslesen kann. In ihn werden die zu verarbeitenden Daten eingelesen und im Verlaufe der Befehlsabarbeitung anfallende Zwischenergebnisse abgespeichert. Der Schreib-Lese-Speicher verliert im Gegensatz zum Festwertspeicher beim Ausfall der Stromversorgung seinen Inhalt. Der Mikroprozessor selbst steuert den Ablauf der Signalverarbeitungsprozesse des Mikrorechners. Er bestimmt, wann und zwischen welchen Teilen des Mikrorechners In-

formationen über die Verbindungsleitungen (Bussystem: Adreß-, Daten- und Steuerbus) übertragen werden. Er entnimmt dem Programmspeicher Befehle, führt sie aus und hat dabei Zugriff zu den Speichern und zur Ein- und Ausgabeeinheit. Innerhalb des Mikroprozessors ist eine Gliederung in Funktionsblöcke vorhanden (vgl. Abb. 8.2.6–3). Das *Steuerwerk (Leitwerk)* entnimmt die Arbeitsanweisungen dem Programmspeicher in der vorgegebenen Reihenfolge und entschlüsselt sie. Dann werden die Speicherplätze (Adressen) aufgesucht, in denen die Daten enthalten sind, mit denen logische oder Rechenoperationen durchgeführt werden. Nach der Übertragung der Daten in das *Rechenwerk* und der Ausführung des Befehls wird die Adresse des nächsten Befehls aufgesucht. Das Rechenwerk wird auch als *Arithmetik-Logik-Einheit* bezeichnet. Es nimmt mit den über den Datenbus zugeführten Daten arithmetische (z. B. Addition) oder logische (z. B. Negation) Verknüpfungen vor. Zum Rechenwerk gehören noch Datenregister und Speicher zur kurzzeitigen Operandenspeicherung.

Mikroprozessoren werden vorwiegend in MOS-Technologie (vgl. 8.1.3.4.) hergestellt. Bei Chipflächen um 30 mm^2 enthalten sie mehrere zehntausend Funktionselemente bei einer Verlustleistung um 1 W.

Die Anwendung von Mikroprozessoren bzw. Mikrorechnern in der Prozeßsteuerung ist außerordentlich vielfältig. Als Auswahl sei angeführt: zur Steuerung von Werkzeugmaschinen, Taktstraßen, Setzmaschinen, Dosierautomaten, Tanksäulenabrechnungen, in Kraftfahrzeugzündsystemen, medizinischen Geräten und bei Patientenüberwachung, zur Steuerung von Wasch- und Nähmaschinen und elektronischen Kameras sowie im wissenschaftlichen Gerätebau und in der Rechentechnik selbst.

9. Physikalische Chemie

Die Physikalische Chemie ist ein Teilgebiet der Chemie, das sich Mitte des 19. Jahrhunderts als selbständige Disziplin, als Grenzwissenschaft zwischen Physik und Chemie herausgebildet hat. In der Folgezeit und besonders in den letzten Jahrzehnten werden Stoffumwandlungen mit den chemischen Elementen als kleinsten Bausteinen (*chemische Reaktionen*) verstärkt mit physikalisch-chemischen Methoden untersucht. Diese lassen sich vielfach unmittelbar auf physikalische Grundlagen zurückführen. Wegen der Vielfalt der Stoffe und der Komplexität chemisch reagierender Systeme werden die experimentellen Verfahren der Physik dabei modifiziert und die theoretischen Modelle mehr oder weniger vereinfacht. Daraus resultieren eigenständige Inhalte und Methoden der Physikalischen Chemie, die heute die gesamte Chemie durchdrungen hat.

Ein Teilgebiet der Physikalischen Chemie beschäftigt sich mit den Wechselwirkungen von elektromagnetischer Strahlung und Materie, die in Abhängigkeit von der Wellenlänge der angewandten Strahlung charakteristisch für die chemische Zusammensetzung und Struktur der Untersuchungsobjekte sind (*Spektroskopie*). Die physikalischen Grundlagen dieser Effekte werden an anderer Stelle behandelt (vgl. 7.8.3.3.). Durch spektroskopische Meßverfahren werden die mikroskopischen Eigenschaften der Atome und Moleküle und ihre Wechselwirkungen experimentell bestimmt, deren Interpretation und Vorausberechnung mit Methoden der *Quantenchemie* erfolgt. Auf quantenchemische Berechnungsverfahren, die sich von der *Schrödinger-Gleichung* ableiten (vgl. 5.6.4.2.), kann hier nicht eingegangen werden.

Die Physikalische Chemie beschäftigt sich vor allem mit der Messung und Beschreibung makroskopischer Eigenschaften von Stoffen, wie Energieinhalt, Temperatur, Reaktionsgeschwindigkeit, sowie elektrischen Größen, die zunehmend auf der Grundlage von mikroskopischen Eigenschaften der Atome und Moleküle unter Anwendung statistischer Methoden beschrieben und interpretiert werden. Waren in der Vergangenheit vor allem gasförmige Systeme, homogene Flüssigkeiten und Festkörper Untersuchungsgegenstand physikalisch-chemischer Methoden, gewinnen in neuerer Zeit Systeme mit Grenzflächen an Bedeutung.

9.1. Chemische Thermodynamik

9.1.1. Thermochemie

Chemische Reaktionen verlaufen bei konstant gehaltener Temperatur entweder unter Aufnahme (*endotherm*) oder Abgabe von Wärme (*exotherm*). Messung oder Berechnung derartiger Reaktionswärmen sind von wesentlicher Bedeutung für die Projektierung und Führung chemischer Prozesse in der Industrie.

9.1.1.1. Messung von Reaktionswärmen. Für die Messung von Reaktionswärmen sind nur solche Reaktionen geeignet, die schnell und praktisch vollständig entsprechend der chemischen Reaktionsgleichung verlaufen, z.B. Verbrennungsreaktionen. Zur Messung von Verbrennungswärmen verwendet man *Bombenkalorimeter* (Abb. 9.1.1–1). Die Probe der zu verbrennenden Substanz wird in das Druckgefäß eingebracht, das mit Sauerstoff mit einem Druck von ≈2 MPa gefüllt wird. Die Verbrennung führt zu einer Temperaturerhöhung des umgebenden Wassermantels, die bestimmt wird. Aus dieser Temperaturerhöhung ergibt sich durch Vergleich mit der bei Verbrennung einer Substanz bekannter Verbrennungswärme gemessenen Temperaturdifferenz

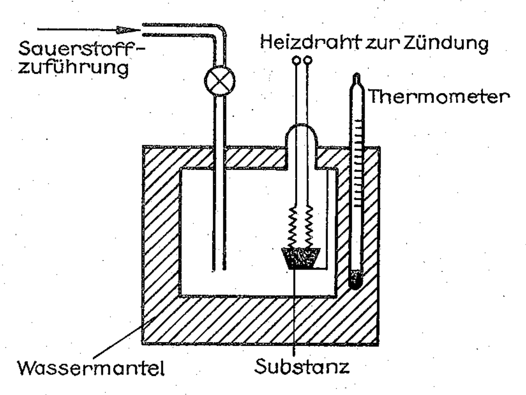


Abb. 9.1.1–1 Bombenkalorimeter zur Bestimmung von Verbrennungswärmen

der gesuchte Wert, der noch bezüglich der Druckänderung während der Reaktion zu korrigieren ist. Nach dem 1. Hauptsatz der Thermodynamik (vgl. 2.6.2.) entspricht die bei konstantem Druck auftretende Reaktionswärme der Änderung der *Enthalpie* ΔH (vgl. 2.6.4.). Die thermochemische Reaktionsgleichung für die Verbrennung von z. B. 1 Mol Benzen schreibt man folgendermaßen:

$$C_6H_6(g) + 15/2\ O_2(g) \rightarrow 6\ CO_2(g) + 3\ H_2O(1);$$
$$H_{298} = -3\,302\ \mathrm{kJmol^{-1}}$$

Durch Indizierung (g = gasförmig, l = flüssig, s = fest) sind die Aggregatzustände der Reaktionspartner sowie die Temperatur (in Kelvin) anzugeben.

9.1.1.2. Hesscher Wärmesatz. Entsprechend dem *Energieerhaltungssatz* gilt, daß die Reaktionswärme unabhängig vom Reaktionsweg, d. h. von der Art und Folge der Teilreaktionen ist. Deshalb können durch geeignete Kombination von thermochemischen Reaktionsgleichungen mit bekannten Werten der Reaktionsenthalpie unbekannte rein rechnerisch indirekt bestimmt werden. Von besonderer Bedeutung ist, daß sich aus den experimentell zugänglichen Verbrennungswärmen (vgl. 9.1.1.1.) die Reaktionswärmen für die Bildung von Verbindungen aus den chemischen Elementen ergeben. Die thermochemischen Reaktionsgleichungen lassen sich hierfür wie algebraische Gleichungen behandeln, z. B. für die Berechnung der Bildungsenthalpie von Benzen:

I $C_6H_6(g) + 15/2\ O_2(g) = 6\ CO_2(g) + 3\ H_2O(l)$;
$\Delta H_I = -3\,302\ \text{kJmol}^{-1}$
II $6\ C(\text{Graphit}) + 6\ O_2(g) = 6\ CO_2(g)$;
$\Delta H_{II} = -2\,361\ \text{kJmol}^{-1}$
III $3\ H_2(g) + 3/2\ O_2(g) = 3\ H_2O(l)$;
$\Delta H_{III} = -858\ \text{kJmol}^{-1}$
$6\ C(\text{Graphit}) + 3\ H_2(g) = C_6H_6(g)$;
$\Delta H_{\text{Benzen}} = -\Delta H_I + \Delta H_{II} + \Delta H_{III} = 83\ \text{kJmol}^{-1}$.

9.1.1.3. Standardbildungsenthalpien. Da nur Enthalpiedifferenzen meßbar sind, kennt die Thermodynamik die Angabe von Absolutwerten der Enthalpien von z. B. 1 Mol eines chemischen Elements oder einer Verbindung nicht. Durch Festlegung des Zustandes der stabilen Modifikationen der Elemente bei 101 kPa und 298 K als Bezugszustand (*Standardzustand*) und willkürliche Zuordnung des Enthalpiewertes Null werden die Reaktionsenthalpien der Bildung chemischer Verbindungen aus den Elementen zu Standardbildungsenthalpien ΔH_f^0, die sich immer auf 1 Mol und den Standardzustand beziehen. Eine große Anzahl derartiger Werte sind in Tabellenwerken zusammengestellt und gestatten die Berechnung beliebiger Reaktionsenthalpien zunächst bei 298 K.

9.1.1.4. Kirchhoffscher Satz. Standardbildungsenthalpien und damit Reaktionsenthalpien ΔH_R sind temperaturabhängig. Der Temperaturkoeffizient der Enthalpie $(\partial H/\partial T)p = c_p$ ist die *Molwärme* bei konstantem Druck, die ihrerseits temperaturabhängig ist. Molwärmen und ihre Temperaturfunktion können durch kalorimetrische Messungen, Temperaturmessungen nach Zufuhr definierter Wärmemengen, experimentell ermittelt werden. Für die Temperaturabhängigkeit von Reaktionsenthalpien gilt:

$$d\Delta H_R/dT = c_p(T)\textstyle\sum_{\text{Produkte}} - c_p(T)\sum_{\text{Reaktanten}}.$$

Bei Kenntnis der Molwärmen von Produkten und Reaktanten einer chemischen Reaktion und ihren Temperaturfunktionen lassen sich deshalb Reaktionsenthalpien durch Integration obiger Gleichung von einer Temperatur auf eine andere umrechnen.

9.1.2. Thermodynamik und chemische Gleichgewichte

In homogenen Reaktionssystemen läuft eine chemische Reaktion $\alpha A + \beta B \rightarrow \chi X + \upsilon Y$ niemals wirklich unter vollständiger Umwandlung der Ausgangsstoffe ab. Wenn nach Ablauf der Reaktion makroskopisch keine chemischen Veränderungen mehr festgestellt werden können, sind zumindest noch geringe Mengen der Ausgangsstoffe vorhanden. Dieser Zustand, der unter definierten äußeren Bedingungen (Druck, Temperatur) durch ein bestimmtes konstantes Verhältnis der Produkte der Konzentration c von End- bzw. Ausgangsstoffen gekennzeichnet ist, heißt *chemisches Gleichgewicht* und wird durch das *Massenwirkungsgesetz* beschrieben: $K = c_X\chi c_Y\upsilon / c_A\alpha c_B\beta$. Zum gleichen Zustand gelangt man, wenn man von den reinen Endstoffen ausgeht. K ist die *Gleichgewichtskonstante*, aus der sich die maximale Ausbeute einer chemischen Reaktion berechnen läßt. Aus dem 2. Hauptsatz (vgl. 2.6.3.), folgt eine wesentliche Grundgleichung der chemischen Thermodynamik: $\Delta G_R^0 = \Delta H_R^0 - T\Delta S_R^0$. Dabei bedeutet auf den Formelumsatz bezogen ΔG_R^0 die *freie Reaktionsenthalpie* (bei reversibler Prozeßführung maximal nutzbare Arbeit), ΔH_R^0 die Reaktionsenthalpie und ΔS_R^0 die Differenz der Entropien von End- und Ausgangsstoffen, jeweils unter Standardbedingungen. Entsprechend dem 3. Hauptsatz (vgl. 2.6.13.) können aus Messungen der Molwärmen Absolutwerte der Entropien bestimmt werden. Nur wenn ΔG_R^0 negativ ist, läuft eine chemische Reaktion freiwillig ab, bis im Gleichgewicht $d\Delta G = 0$ erreicht ist. Der Zusammenhang zwischen Gleichgewichtskonstante und freier Reaktionsenthalpie lautet: $\Delta G_R^0 = -RT\ln K$. ΔG_R^0 wird als freie Standardreaktionsenthalpie bezeichnet. Da man ebenso wie im Falle der Standardbildungsenthalpie freie Standardbildungsenthalpien der chemischen Verbindungen tabelliert hat, wobei die der Elemente im Standardzustand wiederum gleich Null gesetzt werden, lassen sich aus diesen Werten freie Standardreaktionsenthalpien und daraus Gleichgewichtskonstanten berechnen. Wenn sich die Molzahl im Verlauf der Reaktion nicht ändert, gilt bei konstantem Druck für die Temperaturabhängigkeit der Gleichgewichtskonstanten die *van't Hoffsche Reaktionsisobare* $d\ln K/dT = \Delta H_R^0/RT^2$.

Mit den angeführten Beziehungen können solche fundamentalen Fragestellungen der Chemie, ob und in welcher Richtung chemische Reaktionen mit welcher maximalen Ausbeute ablaufen, auf der Grundlage von Messungen der Molwärmen und Reaktionswärmen zumindest prinzipiell beantwortet werden.

Alle bisher angeführten Beziehungen gelten für ideales Verhalten (am besten in verdünnten Gasen experimentell realisierbar), wo zwischenmolekulare Wechselwirkungen vernachlässigt werden können. Rein rechnerisch sind thermodynamische Größen und damit auch Gleichgewichtskonstanten auf der Basis experimentell bestimmter Moleküldaten unter Verwendung des Konzepts der statistischen Thermodynamik grundsätzlich ebenfalls zugänglich.

9.1.3. Mischphasenthermodynamik

9.1.3.1. Zustandsdiagramme. Für die praktische Nutzung chemischer Reaktionen in der Industrie spielen für die Prozeßführung neben Reaktionswärmen und Reaktionsgleichgewichten außerdem Phasengleichgewichte reiner Stoffe bzw. von Stoffgemischen insbesondere bei Trennprozessen eine wichtige Rolle. Unter Phase versteht man den Teil eines Systems, der homogen aufgebaut ist. So gibt es nur

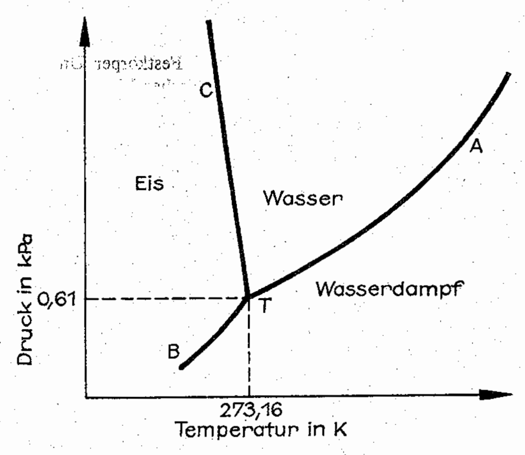


Abb. 9.1.3-1 Zustandsdiagramm des Wassers bei niedrigen Drücken

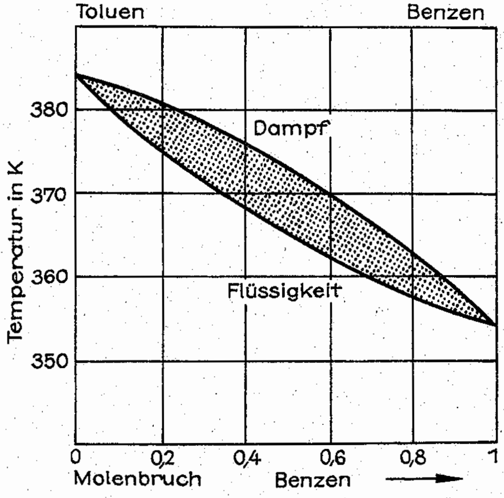


Abb. 9.1.3–2 Siedediagramm des Systems Toluen/Benzen bei 101 kPa. Molenbruch Benzen = Molzahl Benzen/Molzahl Benzen + Molzahl Toluen

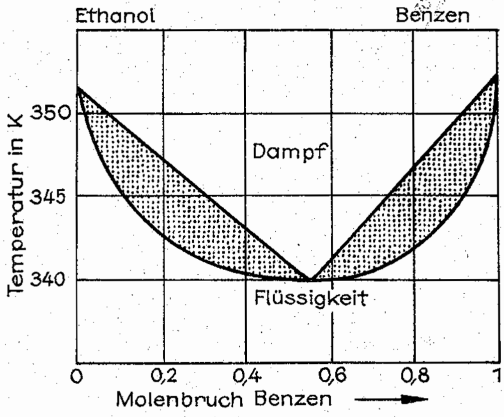


Abb. 9.1.3–3 Siedediagramm des Systems Ethanol/Benzen bei 101 kPa

eine Gasphase. Dagegen können mehrere flüssige und viele feste Phasen nebeneinander existieren. Diese Betrachtungen gelten immer für eingestelltes Gleichgewicht. Eine Phase kann chemisch einheitlich oder aber aus verschiedenen Bestandteilen aufgebaut sein (Mischphase). Die Existenzbedingungen der Phasen eines einkomponentigen Systems lassen sich in einem p, T-Diagramm darstellen (Abb. 9.1.3–1). Darin gibt die Kurve C die Abhängigkeit des Schmelzpunktes vom Druck an. Die Kurven A und B sind die Dampfdruckkurven $p = f(T)$ für flüssiges Wasser bzw. Eis. Für Dampfdruckkurven gilt näherungsweise die *Clausius-Clapeyronsche Gleichung* $d\ln p/dT = \Delta H_{Verdampfung}/(RT^2)$. Beim Tripelpunkt T koexistieren Eis, flüssiges Wasser und Wasserdampf. Für die Trennung flüssiger Systeme durch Destillation haben experimentell durch chemische Analyse von Flüssigkeit und Dampf zu bestimmende Siedediagramme Bedeutung (Abb. 9.1.3–2, 3). Sie geben den Siedepunkt eines Gemisches bei konstantem Druck $d\mu_i = RT d\ln a_i$; Integration liefert $\mu_i = \mu_i^0 + RT\ln a_i$. μ^0 ist das chemische Standardpotential. Der Aktivitätskoeffizient f_i verknüpft Aktivität und Konzentration: $a_i = c_i f_i$. Aktivität und Aktivitätskoeffizient sind vom gewählten Konzentrationsmaß abhängig. Der Standardzustand ist durch Aktivität = eine Konzentrationseinheit gekennzeichnet. Für die Komponenten einer Mischung bzw. das Lösungsmittel einer Lösung entspricht er dem Konzentrationsmaß Molenbruch = Molzahl einer Komponente/Summe der Molzahlen aller Komponenten = Eins, d.h. den reinen Stoffen. Für einen im Unterschuß vorliegenden gelösten Stoff wird der Standardzustand durch das Konzentrationsmaß 1 $mol dm^{-3}$ und Annahme idealen Verhaltens dieser Lösung festgelegt.

Das Massenwirkungsgesetz in thermodynamischer Ableitung für den allgemeinen Fall realer Mischungen lautet: $a_X^{\chi} a_Y^{\upsilon}/a_A^{\alpha} a_B^{\beta} = K_a = e^{-\Delta G_R^0 RT}$. Für binäre Mischungen folgt aus der Gibbs-Duhemschen Gleichung $d\ln a_B = -(c_A/c_B) d\ln a_A$ und $d\ln f_B = -(c_A/c_B) d\ln f_A$.

Phasengleichgewichte wie in 9.1.3.1. beschrieben, sind an die Bedingung geknüpft, daß das chemische Potential jeder Komponente in allen Phasen gleich ist. Darauf aufbauend lassen sich z.B. Aktivitäten des Lösungsmittels in Mischphasen aus Gleichgewichten zwischen Lösungen und reinen Phasen des Lösungsmittels experimentell ermitteln.

9.2. Chemische Kinetik

Während durch eine thermodynamische Betrachtung von Ausgangs- und Endzustand eines Systems festgestellt werden kann, ob und in welcher Richtung eine chemische Reaktion freiwillig abläuft, beschäftigt sich die chemische Kinetik mit der *Reaktionsgeschwindigkeit*, dem Differentialquotienten der auf das Volumen bezogenen Stoffmenge nach der Zeit. Daß thermodynamisch mögliche Reaktionen unter bestimmten Bedingungen nicht oder sehr langsam ablaufen liegt daran, daß bei der Bindungsumordnung zwischen den Atomen im Verlauf einer chemischen Reaktion eine mehr oder weniger hohe Energiebarriere überwunden werden muß, die *Aktivierungsenergie*, die dem reagierenden System zugeführt werden muß. Abb. 9.2.0–1 zeigt schematisch die energetischen Verhältnisse beim reaktiven Stoß zweier Moleküle A und B, der zu X und Y führt. Zunächst steigt die potentielle Energie der beiden wechselwirkenden Moleküle, da zur Lösung von Bindungen Energie aufge-

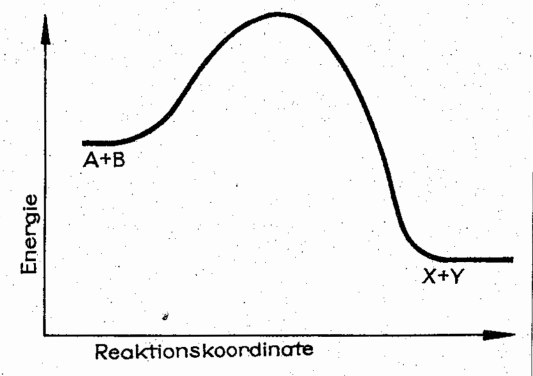


Abb. 9.2.0–1 Energiebarriere im Verlauf der chemischen Reaktion zweier Moleküle

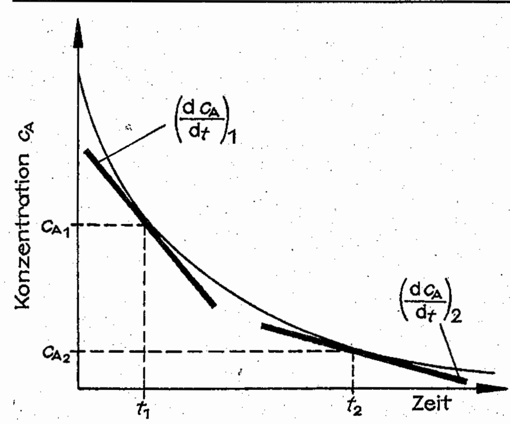


Abb. 9.2.1–1 Konzentrations-Zeit-Diagramm und Reaktionsgeschwindigkeit für den Reaktanten A einer chemischen Reaktion

wandt werden muß. Bei einer bestimmten Konfiguration der beteiligten Atome überwiegt dann die freiwerdende Energie der sich neu bildenden Bindungen und die Energie fällt wieder ab.

9.2.1. Formalkinetik

9.2.1.1. Reaktionsgeschwindigkeit. Für eine Reaktion, deren stöchiometrische Gleichung $\alpha A + \beta B \rightarrow \chi X + \upsilon Y$ lautet, wird die Reaktionsgeschwindigkeit r folgendermaßen definiert:
$r = -(1/\alpha)dc_A/dt = -(1/\beta)dc_B/dt = (1/\chi)dc_X/dt$ $= (1/\upsilon)dc_Y/dt$. Sie ergibt sich aus dem Anstieg der experimentell zu ermittelnden Konzentrations-Zeit-Kurve (Abb. 9.2.1–1). Für konstante Reaktionsbedingungen besteht bezüglich der Konzentrationsabhängigkeit oft folgender Zusammenhang (*Zeitgesetz*): $r = kc_A^a c_B^b$. Die Summe der Exponenten (hier $a + b$) wird als *Reaktionsordnung* bezeichnet. Es kommen auch Zeitgesetze mit komplizierteren Funktionen der Konzentration auf der rechten Seite des Zeitgesetzes (Summen, Brüche) vor, auf die der Begriff der Reaktionsordnung nicht mehr sinnvoll angewandt werden kann. Die *Geschwindigkeitskonstante* k hat den Wert der Reaktionsgeschwindigkeit bei der Konzentrationseinheit. (Man beachte die unterschiedlichen Dimensionen von Geschwindigkeit und Geschwindigkeitskonstante.) Sie ist abhängig von den Reaktionsbedingungen (Temperatur, Lösungsmittel, Katalysator, Größe der Grenzflächen bei Reaktionen in Mehrphasensystemen). Geschwindigkeitskonstanten werden aus experimentellen $r = f(c)$-Daten mittels differentiellem, bzw. aus $c = f(t)$-Daten mittels integriertem Zeitgesetz berechnet. Für die Temperaturabhängigkeit innerhalb des jeweils begrenzten Temperaturbereiches der meisten Meßverfahren der chemischen Kinetik gilt die empirische *Arrhenius-Gleichung* $k = A \exp[-E_a/(RT)]$, A = präexponentieller Faktor, E_a = Aktivierungsenergie. Aufgabe kinetischer Messungen ist also die Ermittlung des Zeitgesetzes, der Geschwindigkeitskonstanten und ihrer Temperaturabhängigkeit. Diese Größen gestatten die Beschreibung des zeitlichen Ablaufs chemischer Reaktionen bei beliebigen Konzentrationen und Temperaturen innerhalb des untersuchten Bereiches.

9.2.1.2. Elementarreaktionen – komplexe Reaktionen. Die formalkinetische Beschreibung geschieht unabhängig davon, ob die Reaktion in mehreren Teilschritten (*komplexe Reaktion*) verläuft, oder die Reaktanten direkt ohne Bildung von Zwischenprodukten zu den Produkten übergehen. (*chemische Elementarreaktion*). Auch im ersten Fall ergeben sich oft einfache Zeitgesetze. Zur gezielten Beeinflussung der Geschwindigkeit chemischer Reaktionen, die meist komplexe Reaktionen sind, muß der *Reaktionsmechanismus*, d.h. Art und Folge der Elementarreaktionen aufgeklärt werden. Wichtigste Aufgabe hierzu ist der chemisch-analytische Nachweis von Zwischenprodukten. Durch Kombination der Zeitgesetze der Elementarreaktionen ergibt sich das Zeitgesetz der komplexen Reaktion. Das entsprechende Differentialgleichungssystem ist nur in den einfachsten Fällen geschlossen lösbar. Die grundlegenden Typen komplexer Reaktionen sind auf den Abb. 9.2.1–2 bis 4 veranschaulicht, in der Realität sind sie oft miteinander verknüpft. Grundsätzliche Vereinfachungen in der formalkinetischen

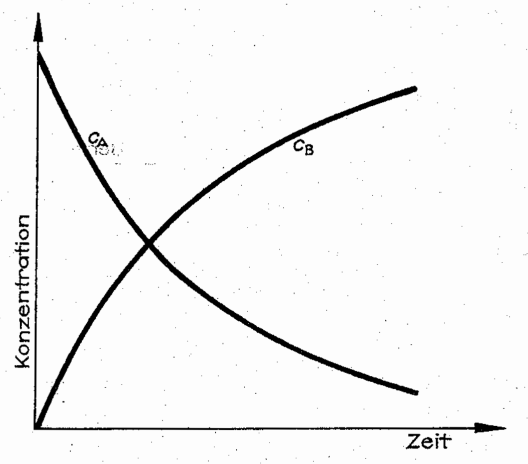


Abb. 9.2.1–2 Konzentrations-Zeit-Diagramm für eine Gleichgewichtsreaktion $A \rightleftarrows B$

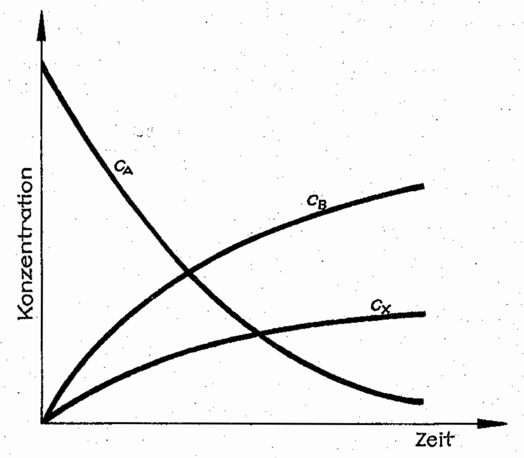


Abb. 9.2.1–3 Konzentrations-Zeit-Diagramm für die Parallelreaktionen $A \rightarrow B$ und $A \rightarrow X$

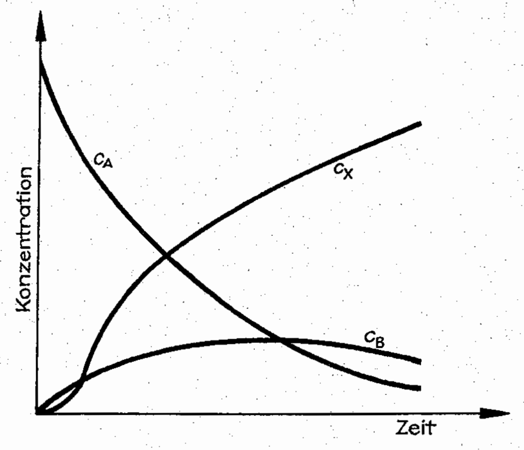


Abb. 9.2.1–4 Konzentrations-Zeit-Diagramm für eine Folgereaktion $A \rightarrow B \rightarrow X$

Beschreibung (Modellierung) komplexer Reaktionen ergeben sich, wenn bei Folgereaktionen die Konzentrationen der instabilen Zwischenprodukte im Verlauf der Reaktion näherungsweise als konstant angesehen werden können (*Bodensteinsches Quasistationaritätsprinzip*). Weiterhin kann oftmals davon ausgegangen werden, daß ein Reaktionsschritt innerhalb einer Folgereaktion im Vergleich zu den anderen mit besonders kleiner Reaktionsgeschwindigkeit abläuft (*geschwindigkeitsbestimmender Schritt*), dessen Zeitgesetz dann als Zeitgesetz der Folgereaktion angesetzt werden kann. Durch diese reaktionskinetische Modellierung komplexer chemischer Reaktionen auf der Grundlage eines sinnvollen Reaktionsmechanismus und experimentell (in einfacheren Reaktionssystemen) bestimmter Geschwindigkeitskonstanten der Elementarreaktionen werden Konzentrations-Zeit-Funktionen von Reaktanten oder Produkten beschrieben oder vorausberechnet. Daraus folgen Hinweise auf den Spielraum der Prozeßparameter in der technischen Reaktionsführung der chemischen Industrie. Grundsätzlich läßt sich ein angenommenes Modell (Reaktionsmechanismus) bei Nichtübereinstimmung zwischen berechneten und experimentell ermittelten Konzentrations-Zeit-Funktionen zwar eindeutig verwerfen, im umgekehrten Fall aber nicht beweisen, sondern nur durch möglichst verschiedenartige, auch nichtkinetische Informationen, weitestgehend experimentell stützen.

9.2.2. Meßmethoden der chemischen Kinetik

Wegen des Zeitbereichs chemischer Reaktionen zwischen 10^{-12} s und $t = \infty$ ist es nicht möglich, Geschwindigkeitskonstanten mit einer einzigen Meßmethode zu bestimmen. Die grundsätzlichen Teilschritte wesentlicher Meßmethoden sind die Herstellung eines definierten Anfangszustandes und danach die Messung der Abhängigkeit der Konzentration von der Zeit. Die Herstellung des Anfangszustandes besteht in der Mischung der Reaktionspartner sowie der Einstellung der Temperatur. Dabei müssen Misch- und Aufheizzeit klein gegenüber der Reaktionszeit sein. Die Messung der Konzentration kann mit allen chemisch-analytischen Verfahren erfolgen, vorausgesetzt, daß sie schnell genug sind und den Reaktionsablauf nicht beeinflussen.

Für Reaktionszeiten $t_R > 1$ min werden statische Meßanordnungen benutzt. Die Reaktionsmischung befindet sich in einem Reaktionsgefäß konstanter Temperatur und in gewissen Zeitabständen wird die Zusammensetzung der Reaktionsmischung chemisch-analytisch bestimmt. Für kürzere Reaktionszeiten bis $t_R > 10^{-3}$ s werden dynamische Verfahren herangezogen, bei denen die Reaktionsmischung durch den Reaktionsraum strömt. Die Reaktionszeit ergibt sich aus dem Quotienten Reaktorvolumen/Strömungsgeschwindigkeit. Die Konzentrationsmessung erfolgt am Ende des Reaktors nach Probennahme oder kontinuierlich durch vor allem spektroskopische Analysenverfahren. In bestimmten flüssigen Reaktionsgemischen läßt sich das Problem der Mischung für sehr schnelle Reaktionen umgehen, indem man vom thermodynamischen Gleichgewicht ausgeht und die Temperatur sehr schnell um einige Grad ändert. Die Konzentrationen der Reaktionspartner stellen sich entsprechend der nach dem Temperatursprung konstanten neuen Temperatur auf die neuen Gleichgewichtswerte mit einer gewissen Verzögerung ein (*Relaxationsmethoden*), die mit schnellen spektroskopischen Methoden bis $t_R > 10^{-6}$ s verfolgt werden können. Falls die zur Überwindung der Aktivierungsbarriere erforderliche Energie durch Lichtquanten aufgebracht werden kann (Fotolyse), lassen sich durch Initierung der Reaktion mit Lichtimpulsen bis minimal ≈ 1 ps (*Laser*, vgl. 4.11.1.5.) die danach ablaufenden Reaktionen kinetisch untersuchen (*Blitzfotolyse*). Daneben wird eine Vielzahl weiterer reaktionskinetischer Meßmethoden genutzt.

9.2.3. Theorie der Reaktionsgeschwindigkeit

Eine theoretische Beschreibung zur Interpretation oder Voraussage von Reaktionsgeschwindigkeiten auf der Basis von Moleküldaten erfaßt die Elementarreaktionen. Sie muß von der Zahl der Stöße (Gasphase) oder Begegnungen (kondensierte Phasen) der reagierenden Atome oder Moleküle ausgehen. So läßt sich für einfachste chemische Reaktionen in der Gasphase der Anteil der reaktiven Stöße durch Behandlung der Bewegung der reagierenden Einzelatome oder -moleküle nach den Gesetzen der klassischen Mechanik (vgl. 2.1.) auf einer quantenmechanisch zugänglichen *Potentialfläche* (vgl. 2.1.3.6.) berechnen. Für fast alle praktisch interessierenden chemischen Reaktionen ist eine derartige Behandlung auch bei Anwendung der modernen Rechentechnik zu kompliziert. Deshalb hat die *Theorie des Übergangszustandes* eine überragende Bedeutung in der Chemie. Darin wird angenommen, daß ein thermodynamisches Gleichgewicht zwischen den Reaktanten und einem Übergangszustand oder „aktivierten Komplex" besteht. Dieser aktivierte Komplex hat die Molekülgeometrie, die dem Energiemaximum auf der Potentialfläche über den sich im Verlauf der reaktiven Wechselwirkung ändernden Atomkoordinaten entspricht (vgl. Abb. 9.2.0–1). Diese Moleküldaten sind näherungsweise aus quantenchemischen Berechnungen zugänglich. Die Reaktionsgeschwindigkeit ist nun proportional der Konzentration des aktivierten Komplexes, die sich aus der Berechnung der Konstanten für das angenommene Gleichgewicht mittels Methoden der statistischen Thermodynamik ergibt. Für die praktische Anwendung der Theorie des Übergangszustandes ist die *Eyring-Gleichung* bedeutsam, $k = (k_B T/h) \cdot \exp(\Delta S^{+}/R) \exp[-\Delta H^{+}/(RT)]$, *wobei* ΔS^{+} = Differenz der Entropien, ΔH^{+} = Differenz der Enthalpien von Reaktanten und akti-

viertem Komplex. Auch wenn für die meisten Elementarreaktionen eine genauere rechnerische Ermittlung dieser Aktivierungsparameter derzeit ebenfalls nicht möglich ist, so lassen sich doch mit qualitativen oder halbquantitativen Abschätzungen über die Veränderung der Atome im Verlauf einer Reaktion Abstufungen und Tendenzen der Aktivierungsparameter innerhalb von Reihen ähnlicher Reaktionen erklären oder voraussagen.

9.3. Elektrochemie

Ein nicht unbedeutender Teil chemisch reagierender Systeme sind wässrige Elektrolytlösungen, d. h. Lösungen von Salzen, Säuren und Basen. Die Untersuchung der elektrischen Leitfähigkeit dieser Systeme, die praktische Bedeutung der Möglichkeit, durch Aufwendung elektrischer Energie chemische Reaktionen auszulösen (*Elektrolyse*) bzw. durch geeignete Anordnungen elektrischen Strom aus dem Ablauf chemischer Reaktionen zu gewinnen (*galvanische Zellen*), trugen wesentlich dazu bei, die physikalische Chemie als selbständige Disziplin innerhalb der Chemie herauszubilden. Die Nutzung elektrochemischer Gesetzmäßigkeiten zur qualitativen und quantitativen chemischen Analyse und die Erzeugung von elektrischem Strom mittels Sonnenenergie sind praktische Aufgabenstellungen der modernen Elektrochemie.

9.3.1. Eigenschaften von Elektrolytlösungen

Ionenkristalle, Säuren und Basen (*Elektrolyte*) lösen sich in Wasser mehr oder weniger gut, weil sie in positive (*Kationen*) und negative (*Anionen*) Ladungsträger (*Ionen*) dissoziieren und Wassermoleküle anlagern, die die Ionen in Lösung stabilisieren. Derartige Elektrolytlösungen, aber auch nichtwässrige Elektrolytlösungen sowie Salzschmelzen leiten den elektrischen Strom. Im Gegensatz zur elektronischen Leitfähigkeit (vgl. 2.4.3.3.) ist die ionische Leitfähigkeit mit einem Stofftransport verbunden. Die Leitfähigkeit L ist der Reziprokwert des elektrischen Widerstandes R, der in einer *Leitfähigkeitsmeßzelle* und mittels einer Wheatstonschen Brückenschaltung (vgl. 2.4.3.) experimentell bestimmt werden kann. Die Messung erfolgt mit Wechselstrom, um Elektrolyse zu verhindern. Es gilt: $L = 1/R = \varkappa A/l$ ($\varkappa$ = spezifische Leitfähigkeit, A = Elektrodenquerschnitt, l = Elektrodenabstand, Maßeinheit für L ist $\Omega^{-1}\text{cm}^{-1}$). Die Zellkonstante A/l wird jedoch nicht durch Ausmessen, sondern durch Bestimmung der Leitfähigkeit einer Kaliumchloridlösung ermittelt, deren spezifische Leitfähigkeit bekannt ist. Zum Vergleich verschiedener Elektrolytlösungen wird auf die Anzahl der Äquivalente N (Molzahl des Elektrolyten geteilt durch die größte Zahl von Ladungen bei Kation oder Anion) je Volumeneinheit (dm^3) bezogen und die *Äquivalentleitfähigkeit* angegeben, $\Lambda = 1\,000\varkappa/N$. Der zu extrapolierende Grenzwert der Äquivalentleitfähigkeit für unendlich verdünnte Lösungen setzt sich additiv aus den Grenzleitfähigkeiten der beteiligten Ionen zusammen $\Lambda^0 = \Lambda^0_+ + \Lambda^0_-$ (*Gesetz der unabhängigen Ionenwanderung*). Elektrolyte mit Äquivalentleitfähigkeiten, für die $\Lambda \sim \sqrt{c}$ gilt, bezeichnet man als *starke Elektrolyte*, die praktisch vollständig in Ionen dissoziiert sind. Für *schwache Elektrolyte*, die nur zu einem geringen Bruchteil in Ionen dissoziiert sind, ergibt sich der Dissoziationsgrad $\alpha = \Lambda/\Lambda^0$. Daraus folgt der Zusammenhang mit der Gleichgewichtskonstanten z. B. für einen 1,1-Elektrolyten (1 Kation, 1 Anion) $K = c\alpha^2/(1-\alpha)$. Da die Ionen immer nur zusammen mit ihren Gegenionen auftreten, können individuelle Aktivitäten bzw. Aktivitätskoeffizienten einzelner Ionensorten auf thermodynamischem Wege nicht bestimmt werden. Deshalb ordnet man beiden Ionensorten denselben mittleren Wert zu und definiert die mittlere Ionenaktivität z. B. eines 1,1-Elektrolyten als das geometrische Mittel der individuellen Ionenaktivitäten, $a^{\pm} = \sqrt{a_+ a_-}$ bzw. $f^{\pm} = \sqrt{f_+ f_-}$. Abweichungen vom idealen Verhalten, d. h. $f \neq 1$ sind auf interionische Wechselwirkungen zurückzuführen, die von den Konzentrationen c_i und den Ladungen z_i aller in der Lösung vorhandenen Ionen abhängig sind. Als empirisches Maß für die Gesamtkonzentration im Hinblick auf die interionischen Wechselwirkungen dient die Ionenstärke

$$I = 1/2 \sum_{i=1}^{n} c_i z_i^2.$$

Die mittleren Aktivitätskoeffizienten starker Elektrolyte lassen sich für geringe Konzentrationen ($\leq 10^{-2}\,\text{mol}\,\text{dm}^{-3}$ für 1,1-Elektrolyte) nach der *Debye-Hückel-Theorie* berechnen. In dieser Theorie wird davon ausgegangen, daß nur Coulombsche Wechselwirkungskräfte (vgl. 2.1.3.3.) wesentlich sind. Es wird die Vorstellung einer jeweils entgegengesetzt geladenen Ionenwolke um jedes Ion zugrundegelegt. Über die Berechnung des Radius dieser Ionenwolke gelangt man zu dem Energieüberschuß, den ein Ion in der Elektrolytlösung gegenüber der rein thermischen Energie hat. Daraus folgt ein Ausdruck für den mittleren Aktivitätskoeffizienten bei 298 K: $f_{\pm} = 0{,}509\,1 z^2 \sqrt{I}$.

9.3.2. Elektromotorische Kräfte

Eine *elektrochemische Zelle* besteht aus zwei Elektroden in einer Elektrolytlösung, die durch einen äußeren Stromkreis miteinander verbunden sind (Abb. 9.3.2–1). Darin ablaufende Reaktionen werden als Redoxreaktionen bezeichnet, weil miteinander gekoppelt Oxidations- (Abgabe von Elektronen) und Reduktionsschritte (Aufnahme von Elektronen) ablaufen. Wenn die an die Zelle gelegte Gleichspannung einen bestimmten Wert (*Zellspannung*) überschreitet, läuft an der Katode die Reaktion $2H^+ + 2e^- \rightarrow H_2$ und an

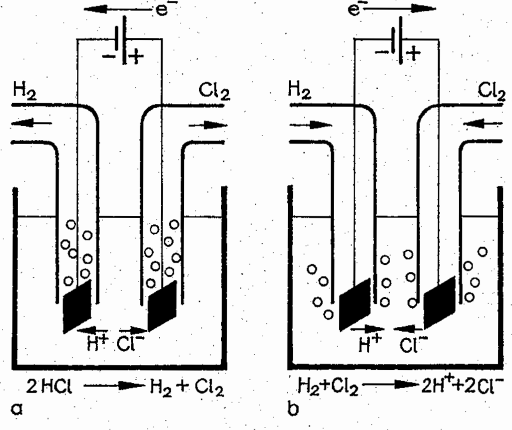


Abb. 9.3.2–1 Elektrochemische Zelle als (a) Elektrolysezelle und (b) galvanische Zelle

der Anode die Reaktion $2\,Cl^- \rightarrow Cl_2 + 2e^-$, d. h. insgesamt die Reaktion $2\,H^+ + 2\,Cl^- \rightarrow H_2 + Cl_2$ ab. Ist die angelegte Spannung kleiner als die Zellspannung, laufen die Elektrodenreaktionen und damit auch die Gesamtreaktion in umgekehrter Richtung (wobei Wasserstoff und Chlor eingeleitet werden müssen). Dabei wird elektrischer Strom in den äußeren Stromkreis abgegeben. Allgemeiner läßt sich der Zusammenhang zwischen oxidierten und reduzierten Formen der Reaktionspartner folgendermaßen formulieren: $ox_1 + red_2 \rightleftarrows red_1 + ox_2$. Chemisches Gleichgewicht herrscht, wenn keinerlei chemische Reaktion mehr vor sich geht, wenn also die angelegte Spannung die Zellspannung gerade kompensiert. Die *elektrochemische Gleichgewichtsbedingung* lautet $\Delta G_R + \Delta E_{elektrisch} = 0$. Werden bei der Reaktion je molaren Formelumsatz N_A-Elektronen ausgetauscht, so beträgt die vom System reversibel geleistete Arbeit $\Delta E_{elektrisch} = N_A e \Delta E$. Mit der Ladungsmenge von 1 Mol Elektronen $F = N_A e = 96\,493\ \mathrm{Cmol^{-1}}$ folgt für die Elektromotorische Kraft (EMK) $\Delta E = -\Delta G/nF$ (n = Zahl der ausgetauschten Elektronen je Atom bzw. Molekül). Durch die Messung der EMK im Gleichgewicht (stromlos) ist es möglich, die freie Reaktionsenthalpie elektrochemischer Reaktionen zu bestimmen (Abb. 9.3.2–2). Für die Konzentrationsabhängigkeit der EMK bei konstantem Druck und konstanter Temperatur gilt die *Nernstsche Gleichung* $\Delta E = -\Delta G^0/nF - RT/nF \ln(a_{ox_2} a_{red_1}/a_{ox_1} a_{red_2})$. Dabei bedeutet $-\Delta G^0/nF = \Delta E^0$ die Standard-EMK, die man messen würde, wenn die Aktivitäten aller Reaktionspartner gleich Eins wären. Man kann nun die EMK einer Zelle als Differenz zweier Elektrodenpotentiale auffassen $\Delta E^0 = E_1^0 - E_2^0$.

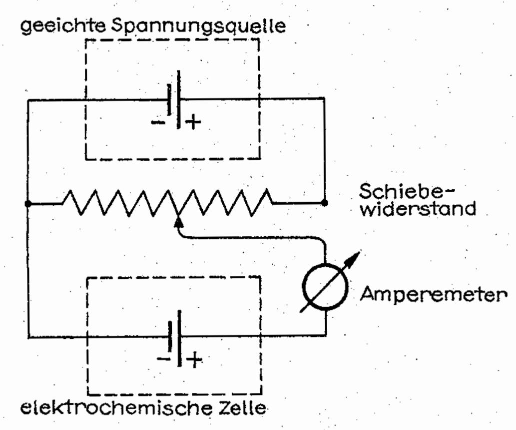


Abb. 9.3.2–2 Prinzip der potentiometrischen Messung der EMK

Da andererseits nur Potentialdifferenzen gemessen werden können, ordnet man der Normalwasserstoffelektrode ($2H^+ + 2e^- \rightarrow H_2$ an Platin bei allen Temperaturen und $a_{H^+} = 1$, $p_{H_2} = 101$ kPa – entspricht der linken Elektrode von Abb. 9.3.2–1) willkürlich das Standardpotential Null zu. Auf diese Weise können *Elektroden-Standardpotentiale* für beliebige Redoxpaare in wässriger Lösung angegeben werden (Tab. 9.3.2–3). Deren Messung erfordert allerdings etwas umfangreichere experimentelle Anordnungen als in Abb. 9.3.2–2 dargestellt. Wegen der schlechten Handhabbarkeit der Normalwasserstoffelektrode verwendet man meist andere Bezugselektroden mit bekanntem Standardpotential, z. B. die Kalomelelektrode $Hg/Hg_2Cl_2/Hg^+$. Ein negatives Standardpotential für die Beispiele der Tabelle bedeutet, daß bei Standardbedingungen an der Vergleichselektrode Wasserstoff entwickelt und das Metall in Lösung gehen würde, ein positives Vorzeichen, daß Wasserstoff als Protonen in Lösung und das Metall an der Elektrode abgeschieden würde.

Durch EMK-Messungen lassen sich thermodynamische Größen, wie freie Standardreaktionsenthalpien, Aktivitäten und Aktivitätskoeffizienten, auf der Grundlage der Nernstschen Gleichung bestimmen, vorausgesetzt, daß geeignete elektrochemische Zellen zusammengestellt werden können.

Von großer Bedeutung ist die Messung der Aktivität der Protonen in wässrigen Lösungen, die für viele chemische Reaktionen eine entscheidende Rolle spielt. Das Ionenprodukt des Wassers $K_W = a_{H_2O} K_a = a_{H^+} a_{OH^-}$ ist eine temperaturabhängige Konstante, aber die Einzelionenaktivitäten lassen sich durch Zugabe von Säuren oder Basen in weiten Grenzen verändern. Zur Auswertung wird die Gleichung $\Delta E = \Delta E^0 - (RT/F)\ln a_{H^+}$ herangezogen, in der ΔE^0 eine für die spezielle Meßanordnung zu ermittelnde Konstante ist. Die Aktivität der Protonen in wässrigen Lösungen wird in der Chemie gewöhnlich als pH-Wert $= -\lg a_{H^+}$ angegeben.

Tab. 9.3.2–3 Elektroden-Standardpotentiale für wäßrigsaure Lösungen bei 298 K

Elektrode	Elektrodenreaktion	E° in V
Li/Li^+	$Li^+ + e^- \rightarrow Li$	−3,05
Al/Al^{+++}	$Al^{+++} + 3e^- \rightarrow Al$	−1,66
Zn/Zn^{++}	$Zn^{++} + 2e^- \rightarrow Zn$	−0,76
Fe/Fe^{++}	$Fe^{++} + 2e^- \rightarrow Fe$	−0,44
$Pt\text{-}H_2/H^+$	$2H^+ + 2e^- \rightarrow H_2$	0,00
Cu/Cu^{++}	$Cu^{++} + 2e^- \rightarrow Cu$	0,34
Ag/Ag^+	$Ag^+ + e^- \rightarrow Ag$	0,80

10. Geophysik

Geophysik ist Physik des Planeten Erde, die Lehre von den physikalischen Zuständen und Prozessen des festen Erdkörpers, der Hydro- und der Atmosphäre, von den Wechselbeziehungen zwischen diesen Geosphären und von den extraterrestrischen Einflüssen auf diese; sie ist somit eine objektorientierte Physik.

10.1. Planet Erde

10.1.1. Astronomische und physikalische Daten der Erde

Die Erde gehört zur Gruppe der inneren Planeten des Sonnensystems (Tab. 10.1.1–1). Sie umläuft die Sonne auf einer elliptischen Bahn mit einer *Exzentrizität* (Entfernung zwischen Mittelpunkt und Brennpunkt der Bahn, geteilt durch die Länge der großen Halbachse) von 0,016 7; größte und kleinste Sonnenentfernung verhalten sich wie 61 : 59. Die mittlere Bahngeschwindigkeit beträgt 29,765 km/s, die Äquatorebene ist 23° 27′ gegen die Ekliptik geneigt.
Neben starken Ähnlichkeiten zu den Nachbarplaneten Venus und Mars in bezug auf Größe, Masse, Dichte und Stoff unterscheidet sich die Erde von diesen durch ihr für biologische Entwicklungen geeignetes Temperaturregime, das sich als Quasigleichgewicht zwischen solarer *Einstrahlung* und terrestrischer *Ausstrahlung* einstellt und das durch terrestrische Eigenschaften, besonders die Existenz einer *Hydrosphäre* mit spezifischer Land-Meer-Verteilung und die selektive Absorption solarer und terrestrischer Strahlung durch die *Atmosphäre* mitbestimmt wird. Wichtige Daten des Planeten Erde sind in Tab. 10.1.1–2 zusammengefaßt.

10.1.2. Physikalische Prozesse der Bildung und Entwicklung

Die zahlreichen Hypothesen über die Entstehung des Sonnensystems verdichten sich in jüngerer Zeit um das Modell der Entstehung „auf kaltem Wege", durch Gravitationsinstabilität im kosmischen Urnebel, die zu einer zentralen Massenakkumulation mit der Bildung der Sonne führte und zu sekundären Akkumulationen mit der Bildung der Planeten.
Die Umsetzung der Gravitationsenergie in Wärme und die Wärmefreisetzung durch Zerfall radioaktiver Substanzen (vor allem der Uranium-, Radium- und Thoriumreihe) hat zu einem teilweisen Aufschmelzen des Planetenkörpers geführt, wobei wahrscheinlich Oberflächentemperaturen in Nähe des Schmelzpunktes der Silikate erreicht worden sind. Resultat dieses Prozesses sind Stoffdifferentiationen, die kugelschalenartigen stofflichen und physikalischen Strukturen des festen Planetenkörpers, die mit Hilfe der *Seismologie* nachgewiesen werden können und über deren Entstehungsprozeß nur erst uneinheitliche Hypothesen existieren. Es gibt Vorstellungen, wonach erst nach abgeschlossener Stoffakkumulation der Erde im Rahmen dieses Aufschmelzprozesses die Stoffdifferentiation eintrat, und andere, die die Bildung des Erdkerns und der Schalenstruk-

Tab. 10.1.0–1 Wichtige Daten der Planeten

	mittlerer Abstand an der Sonne in 10^6 km	Exzentrizität der Bahn	Äquatordurchmesser in 10^3 km	Masse in 10^{24} kg	mittlere Dichte in $g \cdot cm^{-3}$	Fallbeschleunigung an der Oberfläche in $m \cdot s^{-2}$	Rotationsdauer	planetare Albedo	solares Strahlungsdargebot in $W \cdot m^{-2}$
Merkur	57,91	0,205 6	4,876	0,330	5,44	3,82	58 d 15 h	0,06	9 200
Venus	108,21	0,006 8	12,112	4,868	5,23	8,83	242 d 23 h 4 min	0,77	2 700
Erde	149,6	0,016 7	12,756	5,975	5,52	9,806 65	23 h 56 min 4 s	0,30	1 396
Mars	227,9	0,093 4	6,787	0,639	3,95	3,73	24 h 37 min 23 s	0,20	600
Jupiter	778,3	0,048 5	143,650	1 899	1,30	24,61	9 h 50 min	0,67	52
Saturn	1 427	0,055 6	120,670	568	0,68	10,40	10 h 14 min	0,69	15
Uranus	2 870	0,047 2	51,800	87	1,21	8,63	10 h 49 min	0,93	4
Neptun	4 496	0,0086	49,200	103	1,65	11,38	15 h 40 min	0,84	1,5
Pluto	5 946	0,253	≈ 5	≈ 1	2…4	≈ 9	6 d 9 h 17 min	0,14	0,9

Tab. 10.1.0–2 Wichtige Daten der Planeten

Äquatorradius	6 378,388 km
Polradius	6 356,912 km
Abplattung	1 : 297
Oberfläche	510 100 933 km²
Volumen	$1{,}083\,32 \cdot 10^{12}$ km³
Radius einer volumengleichen Kugel	6 371,211 266 km
Bahngeschwindigkeit eines Äquatorpunktes	$465\ m \cdot s^{-1}$
Winkelgeschwindigkeit der Rotation	$7{,}29 \cdot 10^{-5}\ s^{-1}$
mittlere Dichte der Erdkruste	$2{,}60\ g \cdot cm^{-3}$
Hauptträgheitsmoment	$8{,}18 \cdot 10^{37}\ kg \cdot m^2$
magnetisches Moment	$8 \cdot 10^{15}\ Wb \cdot m$

Tab. 10.1.3–1 Erdgeschichtliche Zeittafel

Äon	Ära	Periode	Epoche	Alter in 10^6 a
Phanerozoikum	Känozoikum	Quartär	Holozän	0,015
			Pleistozän	1,5
		Tertiär – Neogen	Pliozän	10
			Miozän	25
		Tertiär – Paläogen	Oligozän	37
			Eozän	58
			Paläozän	67
	Mesozoikum	Kreide	Obere	105
			Untere	137
		Jura	Malm	157
			Dogger	172
			Lias	195
		Trias	Keuper	205
			Muschelkalk	215
			Buntsandstein	225
	Paläozoikum	Perm	Zechstein	240
			Rotliegendes	285
		Karbon	Siles	325
			Dinant	350
		Devon	Oberes	359
			Mittleres	370
			Unteres	405
		Silur		440
		Ordovizium		500
		Kambrium	Oberes	515
			Mittleres	540
			Unteres	570
Riphäikum				1 250
Proterozoikum				2 000
Archaikum				2 870
Katarchaikum				

wahrscheinliches Alter der Erdkruste	4 800 Mill. Jahre
wahrscheinliches Alter der Erde	4 900 Mill. Jahre
wahrscheinliches Alter des Sonnensystems	5 000 Mill. Jahre

turen bereits im Akkumulationsstadium erklären. Als Produkt der Planetenentgasung entstand zunächst eine „*Uratmosphäre*", die im Verlaufe der Planetenevolution mit der Einstellung auf eine tiefere Gleichgewichtstemperatur an der Erdoberfläche unter Wasserdampfkondensation zur Entstehung einer *Hydrosphäre* führte und die sich auf der Erde in Wechselwirkung mit der biologischen Evolution in die Stickstoff-Sauerstoff-Atmosphäre der Gegenwart umbildete. Über die chemische Zusammensetzung der *Uratmosphäre* gibt es unterschiedliche Auffassungen. Nach neueren Erkenntnissen ist es sehr wahrscheinlich, daß die Erduratmosphäre ähnliche Zusammensetzung hatte wie die Gegenwartsatmosphären der Nachbarplaneten Venus und Mars und im wesentlichen aus Kohlendioxid (CO_2) und weiteren geringen Relativanteilen von Methan (CH_4), Ammoniak (NH_3), Wasser (H_2O), ... bestand, wobei der größte Teil des exhalierten Wassers zur *Hydrosphäre* kondensierte. Die Umbildung in die N_2-O_2-Atmosphäre der Gegenwart erfolgte über die Fotosynthese von Kohlehydraten aus CO_2 und H_2O über die Grünpflanzen mit der Freisetzung von Sauerstoff. Der abiologische Bildungsprozeß über die Fotodissoziation des Wassers durch solares Ultraviolett scheidet für die Bildung des atmosphärischen Sauerstoffes aus, weil Sauerstoff im dissoziierenden Strahlungsbereich selbst starker Absorber ist und seinen eigenen Freisetzungsprozeß damit unterbindet. Der molekular und in Verbindungen exhalierte stabile Stickstoff bleibt fast völlig in der Atmosphäre.

10.1.3. Erdalter und Altersbestimmung

Die historische Geologie hat sich eine relative *Zeitskala* in der Ablagerungsfolge von Sedimenten und der in ihnen konservierten Reste von Pflanzen und Tieren aufgebaut (Stratigrafie) (Tab. 10.1.3–1).
Neue, physikalische Methoden einer absoluten *Geochronologie* nutzen den Zerfall der Atomkerne bestimmter Elemente. Unter Voraussetzung bekannter Halbwertzeiten ermöglichen Messungen des Verhältnisses zwischen Tochter- und Ausgangselement eine Altersbestimmung. Für die Bestimmung geringer Alter nutzt man vorzugsweise radioaktive Isotope, die sich unter Wirkung energiereicher Neutronen der *kosmischen Strahlung* in der oberen Atmosphäre bilden, besonders Tritium (^3H), Halbwertzeit 12,4 Jahre. Sie geraten in den Stoffaustausch mit der Erdoberfläche und lassen z. B. das Grundwasseralter seit dem letzten Kontakt mit der Atmosphäre bestimmen. Besonders bewährt hat sich das Kohlenstoffisotop ^{14}C ($t_{1/2}$ = 5 600 Jahre), das unter β-Strahlung zu ^{14}N zerfällt und sich zur Altersbestimmung von kohlenstoffhaltigen Substanzen seit deren letzter Aufnahme von CO_2 aus der Atmosphäre eignet. Zur Datierung weit zurückliegender geologischer Prozesse sind grundsätzlich Elemente mit sehr langsamen Zerfall geeignet. Als Fehler in dieser Methode wirken Verluste von Tochterprodukten durch Löslichkeit und Diffusion, wie es z. B. bei reinen Uraniummineralen der Fall ist. Dagegen haben sich Minerale bewährt, in denen Uranium oder Thorium als Nebenkomponenten vorkommen, etwa die Zircone oder Minerale der seltenen Erden.
Der Zerfall des ^{40}K und des ^{87}Rb ist zu Altersbestimmungen

vielseitig anwendbar, da diese im Eruptivgestein in hinreichenden Konzentrationen vorkommen. Allerdings müssen auch hier Verlustmöglichkeiten (des aus ^{40}K austretenden ^{40}Ar und des aus ^{87}Rb entstehenden ^{87}Sr) kritisch berücksichtigt werden.
In den ältesten Gesteinen der Kerne heutiger Kontinente können auf diese Weise Minerale von mehr als $3{,}6 \cdot 10^9$ Jahren festgelegt werden; offenbar war vorher die Erdkruste zu instabil, so daß keine Zeugen früherer Zeit nachweisbar sind.
Untersuchungen von Steinmeteoriten mit kernphysikalischen Methoden ergeben übereinstimmend mit dem Erdalter $\approx 4{,}5 \cdot 10^9$ Jahre.

10.1.4. Die Geosphären

10.1.4.1. Der feste Erdkörper. Informationen über das Erdinnere (mit Ausnahme der obersten, durch Bergbau und Bohrungen erschlossenen Schicht) liegen nur über die Ausbreitung von *Erdbebenwellen* (*Seismologie*) vor, die einen schalenförmigen Aufbau des festen Erdkörpers mit drei Hauptgliedern ergeben: *Erdkruste*, *Erdmantel*, *Erdkern*.
Durch geophysikalische, geologische, geochemische und petrologische Forschungen ist die Erdkruste naturgemäß am besten bekannt. Nach den Dichten und den beobachteten Geschwindigkeiten seismischer Wellen wird das charakteristische Material in oberer und unterer Kruste durch Granit und Gabbro beschrieben, im oberen Mantel durch Peridotit. In der unter den Ozeanen dünneren Kruste fehlt die „Granitschicht". Der Übergang von der Kruste zum Mantel wird durch die *„Mohorovičić-Diskontinuität"*, eine Zone schroffer Geschwindigkeitsänderung der elastischen Wellen, gebildet. Zwischen den Schichten der Kruste liegt die weniger ausgeprägte *„Conrad-Diskontinuität"*. Die Strukturen und Gebilde der Kruste, Faltengebirge, Tiefseegräben usw., mit mannigfaltigen Erscheinungen, wie Erdbeben und Vulkanismus, sind wesentliche gemeinsame Untersuchungsobjekte von Geophysik und Geologie.
Im oberen *Erdmantel* besteht in der *„Gutenberg-Schicht"* eine Region verringerter Wellengeschwindigkeit. Über Tiefherdbeben gibt es Anzeichen von Abweichungen vom kugelsymmetrischen Aufbau des Erdmantels. Offenbar liegt in Konvektionsströmen des oberen Erdmantels die Triebkraft zu *Spreizbewegungen* der Tiefseeböden mit *Drift der Kontinente*.
Transversale Erdbebenwellen dringen in den *Erdkern* nicht ein; daraus wird gefolgert, daß zumindest der äußere Kern in nicht festem Zustand ist.
Neben dieser Gliederung des Schalenaufbaues nach geophysikalischen Gesichtspunkten (vgl. 10.2.5.2.) gibt es Einteilungen nach geochemischen Merkmalen: Die *Lithosphäre* (Gesteinssphäre) ist der aus Kruste und oberem Mantel gebildete, bis ≈ 1200 km Tiefe reichende, im wesentlichen aus silikatischem Material bestehende Teil des Erdkörpers. Innerhalb dieser Lithosphäre wandern die ≈ 100 km mächtigen Großschollen als Platten auf plastischer Unterlage, der *Asthenosphäre*, die bis ≈ 300 km Tiefe reicht. Unter der Lithosphäre liegt die *Chalkosphäre* (1200 bis 2900 km Tiefe, in der hauptsächlich Schwermetalle vermutet werden. Schließlich wird der Erdkern als *Siderosphäre* bezeichnet, in der Eisen und Platinmetalle angesetzt werden.
Die chemischen Elemente der Kruste sind in ihrer Häufigkeitsfolge in Tab. 10.1.4–1 enthalten.

10.1.4.2. Hydrosphäre. Zur *Hydrosphäre* gehören sämtliche Wasserreservoire an der Erdoberfläche und in der oberen Erdkruste, die Meere, Binnenseen, Flüsse, das Boden- und Grundwasser sowie das *Eis* der Meere, der Gebirgsgletscher und des Frostbodens. Die Gesamtheit des Wassereises wird auch unter dem Begriff *Kryosphäre* zusammengefaßt.
Mit $362 \cdot 10^6\,\text{km}^2$ Fläche bedecken die Meere 71 % der Erdoberfläche, 61 % auf der Nord- und 81 % auf der Südhalbkugel. Die irdische Gesamtwassermasse ($1350 \cdot 10^6\,\text{km}^3$) bildet zum größten Teil die Meere ($1310 \cdot 10^6\,\text{km}^3 \mathrel{\hat{=}} 97{,}2\,\%$); im Eis sind $28{,}5 \cdot 10^6\,\text{km}^3 \mathrel{\hat{=}} 2{,}15\,\%$ festgelegt und nur 0,65 % ($\approx 8{,}3 \cdot 10^6\,\text{km}^3$) sind Süßwasser. Der atmosphärische Wasserdampf der Atmosphäre beträgt nur 0,001 % des irdischen Wassers.
Die *hypsometrische Kurve* der festen Erdoberfläche (Abb. 10.1.4–2) veranschaulicht die Höhenunterschiede. Sie ergibt eine mittlere Meerestiefe von 3795 m mit maximalen Tiefen von über 11000 m (Witjas-Tiefe im Marianengraben 11034 m). Die mittlere Höhe des Festlandes beträgt 875 m. Die hypsometrische Kurve läßt topografische Großformen des Meeresbodens erkennen: *Schelfe*, Gürtel bis 200 m tiefer Flachseen um die Kontinente über meist sehr mächtigen Sedimenten; sie gehen über in *Kontinentalabhänge*, die den eigentlichen Rand der Kontinentblöcke bilden, bis zu den *Tiefseebecken* (mit Rücken und Schwellen), die ein relativ ausgeglichenes Relief aufweisen, aber von untermeerischen Gebirgen durchzogen werden. *Tiefseegräben* befinden sich meist an den Rändern von Ozeanen. Sie sind häufig Zonen hoher seismischer Aktivität und gleichzeitig Gebiete negativer Schwereanomalien. Dies deutet auf gegenwärtige *Geosynklinalbildungen* hin, in denen ozeanische Kruste unter Kontinentblöcke taucht.
Der mittlere *Salzgehalt des Meerwassers* beträgt 3,5 %, zum größten Teil (88,6 %) Chlorverbindungen (Kochsalz, Magnesiumchlorid u. a.), außerdem schwefelsaure Salze (10,8 %) und kohlensaure Salze (0,6 %). Der Ursprung des Meersalzes liegt in Freisetzungen von Metallionen der magmatischen Gesteine durch Verwitterung und in sauren vulkanischen Exhalationen, die insgesamt zu mannigfaltigen chemischen Verbindungen führen.
Mit einer mittleren Eisdicke von 2 bis 2,5 m bedeckt Meereis die Schelfe der Polarmeere. Es driftet als *Treibeis* in niedere Breiten. Beim Kalben der *Festlandsgletscher* entstehen *Eisberge*; obwohl nur 1/8 des Eises oberhalb der Meeresoberfläche bleibt, erreichen arktische Eisberge Höhen bis 100 m über Wasser. Auch die antarktischen Tafeleisberge haben Flankenhöhen bis zu 50 m. Etwa $15{,}8 \cdot 10^6\,\text{km}^2$ der Festlandsfläche (über 10 %) sind eisbedeckt. Das Gletschervolumen der Antarktis von $28 \cdot 10^6\,\text{km}^3$ umfaßt 91 % des irdischen Gletschervolumens, das Grönlandeis 8 %. Die Gletscher bilden sich aus dem *Firn*, dem mehrjährigen körnigen

Tab. 10.1.4–1 Elementeverteilung in der Erdkruste

Element	Masse-%
O	46,71
Si	27,69
Al	8,07
Fe	5,05
Ca	3,65
Na	2,75
K	2,58
Mg	2,08
H	0,14

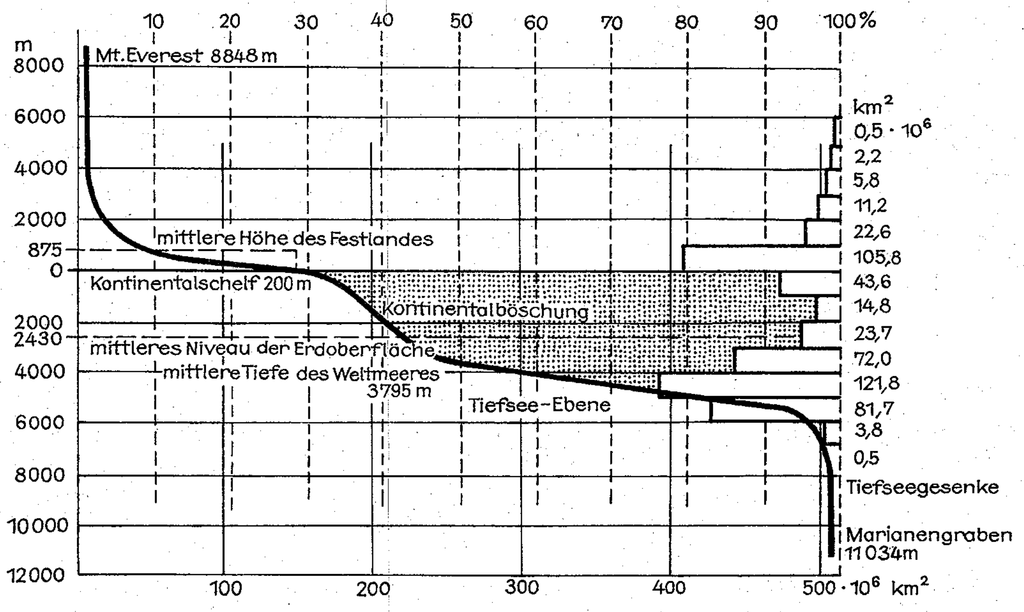


Abb. 10.1.4–2 Hypsometrische Kurve der festen Erdoberfläche

Schnee, durch Verdichten zu *Firneis* und weiter zu *Gletschereis*.

10.1.4.3. Atmosphäre. Die Gashülle der Erde mit einer Gesamtmasse von $5{,}1 \cdot 10^{15}$ t ist ein Gemisch aus verschiedenen Gasen (Tab. 10.1.4–3), vorwiegend molekularer Stickstoff und Sauerstoff, die im Bereich der *Homosphäre* (bis ≈ 80 km Höhe) ein konstantes Mischungsverhältnis bilden, in der darüberliegenden *Heterosphäre* eine Schichtstruktur aufweisen (Abb. 10.1.4–4). Einige Stoffe sind in raumzeitlich variablen Anteilen in der Atmosphäre vorhanden. Hierzu gehören besonders das Kohlendioxid und der Wasserdampf, sowie Spurengase, flüssige und feste Beimengungen (*Aerosol*) aus natürlichen und anthropogenen Quellen.

Tab. 10.1.4–3 Zusammensetzung der Atmosphäre (Homosphäre)

Stoff	Vol.-%	Gesamtmasse in t
Stickstoff	78,09	$3{,}86 \cdot 10^{15}$
Sauerstoff	20,95	$1{,}18 \cdot 10^{15}$
Argon	0,93	$6{,}55 \cdot 10^{13}$
Wasser	variabel	10^{13}
Kohlendioxid	0,03	$2{,}33 \cdot 10^{12}$
Neon	$1{,}8 \cdot 10^{-3}$	$6{,}4 \cdot 10^{10}$
Helium	$5{,}2 \cdot 10^{-4}$	$3{,}7 \cdot 10^{9}$
Krypton	$1{,}5 \cdot 10^{-4}$	$1{,}5 \cdot 10^{10}$
Methan	$1{,}4 \cdot 10^{-4}$	$3{,}9 \cdot 10^{9}$
Wasserstoff	$5 \cdot 10^{-5}$	$2{,}3 \cdot 10^{8}$
Xenon	$8 \cdot 10^{-6}$	$2{,}5 \cdot 10^{9}$
Ozon	$\approx 10^{-8}$	$\approx 10^{9}$
Radon	$\approx 10^{-18}$	0,08...22
weitere Spurengase und Aerosol		

Nach physikalischen Eigenschaften, besonders nach dem vertikalen Temperaturverlauf, wird die Atmosphäre in Stockwerke gegliedert. Die in ihrer Gesamtheit je nach Zeitmaßstab als *Wetter*, *Witterung* und *Klima* bezeichneten dynamischen und thermodynamischen Prozesse bzw. Zustände spielen sich im wesentlichen in der *Troposphäre* und der unteren *Stratosphäre* ab.

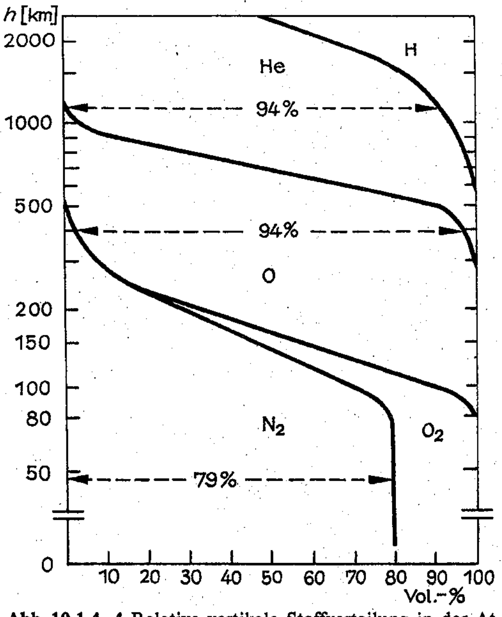


Abb. 10.1.4–4 Relative vertikale Stoffverteilung in der Atmosphäre

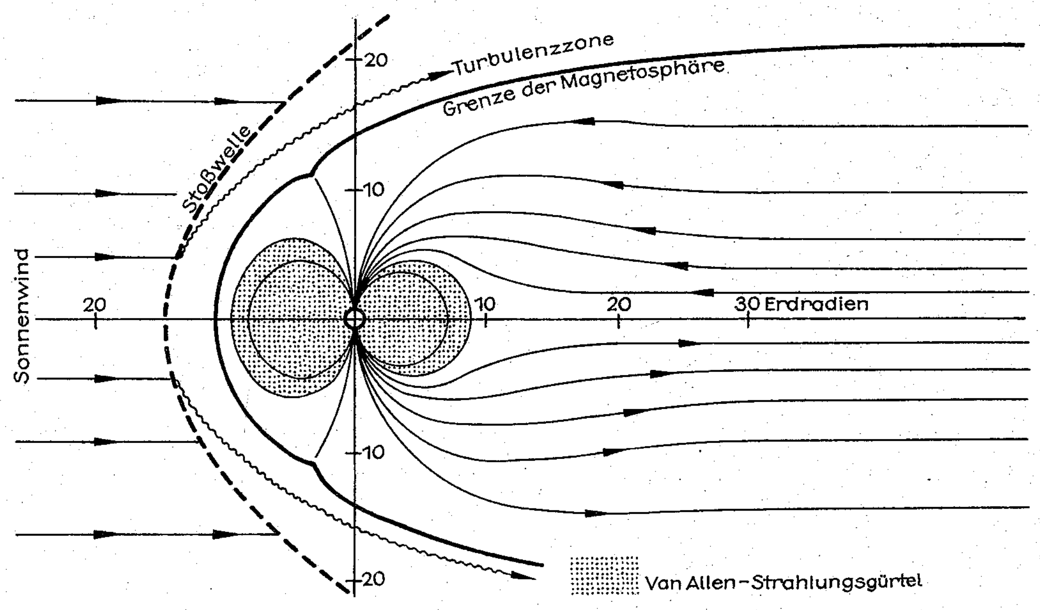


Abb. 10.1.4–5 Schema der Magnetosphäre

Die relative Ozonanreicherung (O_3) in der mittleren Stratosphäre filtert das kurzwellige solare UV ($\lambda < 290$ nm) so vollständig aus, daß diese biologisch schädliche Strahlung nicht bis zur Erdoberfläche dringt.
In der *Heterosphäre* werden molekularer Stickstoff und Sauerstoff durch die *Sonnenstrahlung* des UV- und des Röntgen-Bereiches sowie durch *Korpuskularstrahlung* ionisiert. Es bilden sich raumzeitlich variable Konzentrationen freier Elektronen in bevorzugten Niveaus mit Einfluß auf Absorption und Refraktion von Radiowellen aus.
Im Schwerefeld der Erde wird die Atmosphäre beständig festgehalten und bildet ein mit der Erde rotierendes System mit spezifischen inneren physikalischen Prozessen (vgl. 10.4.). Oberhalb 500 bis 600 km Höhe sind die mittleren freien Weglängen so groß, daß neutrale Luftmoleküle, deren thermische Geschwindigkeit größer als die Fluchtgeschwindigkeit von der Erde (11 km/s) ist, aus dem Schwerefeld der Erde entweichen können (*Exosphäre*).
Zwischen der Atmosphäre und ihrer festen und flüssigen Unterlage vollzieht sich ein sehr vielseitiger Stoff- und Energieaustausch, der zusammen mit extraterrestrischen Einflüssen, besonders mit der Sonnenstrahlung, die natürlichen Umweltbedingungen an der Erdoberfläche bestimmt.

10.1.4.4. Magnetosphäre. In der *Exosphäre* liegt die obere Grenze einer neutralen Atmosphäre. In großen Höhen werden jedoch Luftmoleküle und -atome ionisiert und geraten unter den Einfluß des *Erdmagnetfeldes*. Geladene Teilchen bewegen sich deshalb schraubenförmig um magnetische Feldlinien und können die Erde nicht verlassen. Diese so gebundenen Teilchen bilden die *Magnetosphäre*.
Der „*Sonnenwind*", der von der Sonne kommende Partikelstrom, gibt der Magnetosphäre eine schweifartig von der Sonne weggerichtete Form (Abb. 10.1.4–5). Die *Magnetopause* bildet die äußere Grenze des Einflußbereiches des Planeten Erde.

10.1.5. Energiezufuhr von der Sonne

Vom außerirdischen Raum erhält der Planet Erde Energie zugeführt in Form *elektromagnetischer Wellen* (von Röntgen-Strahlen über Licht bis in das Gebiet der Radiowellen) und in Form von Teilchenströmen (*Korpuskularstrahlung*). Beide Energieübertragungsformen treten mit der Erdatmosphäre in Wechselwirkung und übertragen ihre Energie in vielfältigen Prozessen an diese und an die Erdoberfläche (Kontinente und Ozeane).

10.1.5.1. Energiezufuhr über elektromagnetische Wellen. Wegen der hohen *Strahlungsleistung der Sonne* ($3{,}86 \cdot 10^{20}$ MW) und der relativ kleinen Entfernung zur Erde können alle sonstigen Energieströme von außerirdischen Quellen wegen sehr viel geringerer Größenordnungen für energetische Bilanzierungen unberücksichtigt bleiben.
Der von der Sonne ausgehende Strahlungsstrom trifft den Planeten Erde mit einer Strahlungsdichte S_0, die als *Solarkonstante* bezeichnet wird und die gleich der extraterrestrischen (außerhalb der Erdatmosphäre) senkrecht zur Strahlungsrichtung auftreffenden Bestrahlungsstärke E_0 ist. Der mittlere (standardisierte) Wert der Solarkonstanten beträgt $S_0 = 1{,}396$ kW/m². Infolge des im Jahresverlauf veränderlichen Abstandes Sonne–Erde schwankt dieser Wert zwischen 1,44 kW/m² (Ende Dezember) und 1,35 kW/m² (Ende Juni). Außerdem variiert der Wert in mehreren Perioden sehr unterschiedlicher Längen um einige Zehntel Prozent. Nach Meßergebnissen der jüngeren Vergangenheit liegt der genaue Mittelwert S_0 zwischen 1,368 und 1,377 kW/m².
Die solare Strahlungsenergie stammt im wesentlichen aus der Fotosphäre der Sonne, einer ≈ 400 km dicken Schicht an der Sonnenoberfläche. Diese Schicht bestimmt die generelle Energieverteilung über die Wellenlängen des *Spektrums der Sonnenstrahlung*. In erster Näherung kann die Sonne als „schwarzer Strahler" aufgefaßt werden, der nach dem Planckschen Gesetz emittiert. Die Temperatur der Fo-

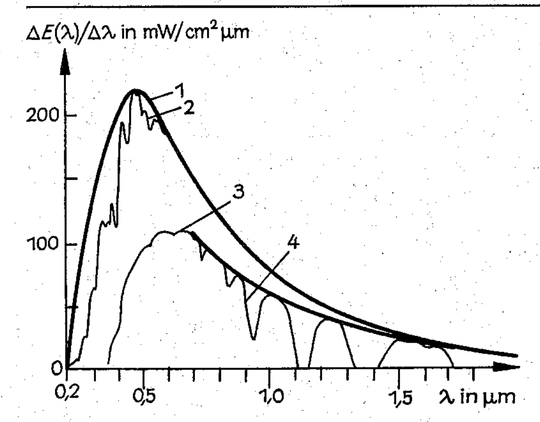


Abb. 10.1.5–1 Spektrum der Sonnenstrahlung (1.: Spektrum eines schwarzen Körpers der effektiven Strahlungstemperatur der Sonne, 5 770 K; 2.: extraterrestrisches Sonnenspektrum; 3.: Sonnenstrahlung (Bestrahlungsstärke) an der Erdoberfläche durch Dunst- und Rayleigh-Extinktion der Atmosphäre bei einer Sonnenhöhe von 24°; 4.: wie 3. zuzüglich der Wasserdampfextinktion)

tosphäre führt zu einer spektralen Strahlungsenergieverteilung (Abb. 10.1.5–1), deren Maximum bei 450 bis 480 nm Wellenlänge liegt, und die durch Absorptions- und Emissionsbanden der Sonnenatmosphäre (solarer Teil der Fraunhoferschen Linien) überlagert wird.

Neben diesem sehr beständigen Strahlungsstrom sendet die Sonne variierende Anteile elektromagnetischer Wellen im Bereich der (sehr kurzen) *Röntgen-Strahlung* und der *Radiowellen* aus, die mit speziellen Sonnenaktivitäten in Verbindung stehen. Diese Strahlungen sind als Energieträger bedeutungslos.

10.1.5.2. Energiezufuhr über Korpuskularstrahlung. Der Energietransport über Materieteilchen erfolgt durch den primären Teilchenstrom (*Korpuskularstrahlung*), der durch Wechselwirkung mit den atmosphärischen Molekülen sekundäre „*Teilchenschauer*" erzeugt. Der primäre Teilchenstrom setzt sich aus der *kosmischen Strahlung* und dem von der Sonne stammenden *Sonnenwind* zusammen. Die kosmische Strahlung besteht vorwiegend aus schnellen Atomkernen und Elektronen, die aus dem Weltraum in die Atmosphäre eindringen. Sie stammen vermutlich aus fernen Galaxien und heben sich durch ihre hohe kinetische Energie von der übrigen kosmischen Materie ab. Die ionisierten Atome der kosmischen Materie haben eine solche Beschleunigung erfahren, daß sie sich nahezu mit Lichtgeschwindigkeit fortbewegen. Die Zusammensetzung der kosmischen Materie bezüglich Energieverteilung und Kernspektrum ist in Abb. 10.1.5–2 dargestellt. Die Energiestromdichte der primären Teilchenstrahlung beträgt $\approx 1\,400\ \mathrm{MeV/cm^2 s}$.

Der Sonnenwind besteht vorwiegend aus Protonen und Elektronen und stellt die expandierende Sonnenkorona dar. Neben einem relativ gleichmäßigen Anteil verursachen Eruptionen auf der Sonne (*Protuberanzen, Solarflares*) wolkenartige Ausbrüche von Teilchen, die sich mit mittleren Geschwindigkeiten von 1 bis $2 \cdot 10^3$ km/s bewegen. Bei chromosphärischen Eruptionen (Solarflares) können Anfangsgeschwindigkeiten bis zur halben Lichtgeschwindigkeit auftreten. Solche Teilchenschauer verursachen auf der Erde *magnetische Stürme* und durch Einschuß geladener Teilchen in das Magnetfeld und ihre Ablenkung zu den Polargebieten *Polarlichter*.

10.1.6. Energiehaushalt des Planeten Erde

Der Energiehaushalt des Planeten Erde setzt sich aus dem in der Atmosphäre und an der Erdoberfläche absorbierten und in Wärme umgesetzten Teil der solaren Einstrahlung E_0 und der terrestrischen Ausstrahlung E_p zusammen. Der von Atmosphäre und Erdoberfläche in den Weltraum reflektierte oder rückgestreute Solarstrahlungsanteil, die *Albedo* der Erde A_p, beträgt im planetaren Mittel 30 %. Die Energieabstrahlung der Erde liegt wegen der tiefen effektiven Strahlungstemperatur von 253 K im thermischen Infrarotbereich mit einer maximalen Energieabstrahlung im Gebiet von 3 bis 50 μm Wellenlänge.

Im Mittel über die gesamte Planetenoberfläche beträgt das solare Strahlungsdargebot $E_0 = S_0/4 = 0{,}344\ \mathrm{kW/m^2}$. Der *planetare Energiehaushalt* ergibt sich somit aus der Summe $E_0(1 - A_p) - E_p = Q$, wenn der Erdwärmestrom wegen sei-

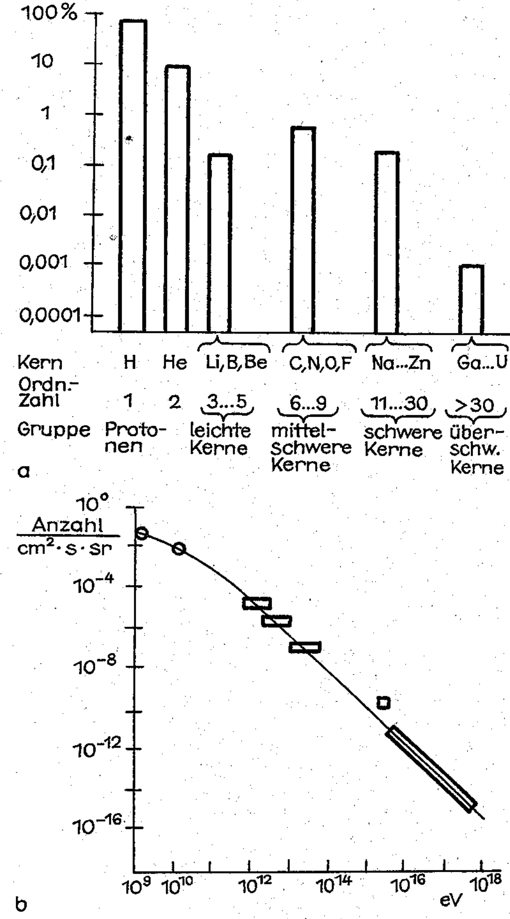


Abb. 10.1.5–2 (a) Kernspektrum und (b) Energieverteilung der kosmischen Strahlung

ner relativ geringen Größe hierbei unberücksichtigt bleibt. Im globalen Mittel und über längere Zeit ($\geqq$ 1 Jahr) stehen Ein- und Ausstrahlung im Gleichgewicht, es ist also $Q = 0$, und es herrscht ein beständiges Temperaturregime. Der planetare Energiehaushalt bestimmt wesentlich den *Wärmehaushalt* an der Erdoberfläche und mit ihm Wetter und Klima (vgl. 10.4.1.).

10.2. Physik des festen Erdkörpers

10.2.1. Figur der Erde

10.2.1.1. Schwerefeld. Die physische Erdoberfläche mit ihren vielgestaltigen Formen ist mit einfachen mathematischen Hilfsmitteln nicht geschlossen darstellbar. Zur Beschreibung von Größe und Form der Erde und ihrer physikalischen Eigenschaften bedient man sich Näherungen, die entweder geometrisch als *Ellipsoid* oder physikalisch als *Geoid* definiert sind. Als Bezugsoberfläche für das Geoid verwendet man ein *Rotationsellipsoid* im Meeresniveau, das deshalb auch *Referenzellipsoid* heißt. Als *„Normalerde"* wird ein Ellipsoid bezeichnet, das die gleiche Rotation und die gleiche Masse wie die wahre Erde hat (Abb. 10.2.1–1).

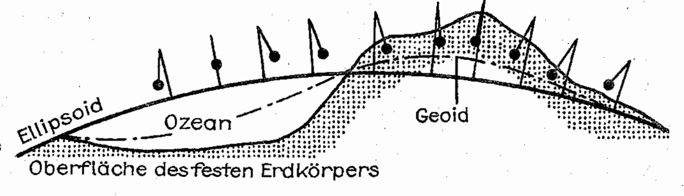


Abb. 10.2.1–1 Schematische Darstellung der physischen Erdoberfläche mit Angabe der Lotrichtungen

Die Erdbeschleunigung entsteht als Resultat der Gravitation und der (breitenabhängigen) Zentrifugalbeschleunigung. Die auf das Meeresniveau des Erdellipsoids bezogene *Normalschwere* wird mit der *internationalen Schwereformel* ausgedrückt:

$$g = 978{,}049\,(1 + 0{,}005\,288\,4 \sin^2\varphi - 0{,}000\,005\,9 \sin^2 2\varphi)\,\mathrm{cm/s^2}$$

(φ = geografische Breite).

Die Änderung der Erdbeschleunigung für Punkte ober- und unterhalb des Meeresniveaus ($\triangleq$ *Freiluftkorrektion*) lautet: $\partial g/\partial h = (0{,}308\,78 + 0{,}000\,43 \sin^2\varphi)\,10^{-3}\,\mathrm{cm/s^2}$. Durch unterschiedliche Masseverteilungen im realen Erdkörper treten regionale *Schwereanomalien* auf.

10.2.1.2. Erdrotation. Die *Rotation der Erde* um ihre Achse ist nicht gleichförmig, und die Achse führt unter dem Einfluß der Gezeitenwirkung die Bewegung eines nicht ganz starren Kreisels aus. Der als *Präzession* bezeichneten Bewegung, deren Periode 25 800 Jahre beträgt, überlagern sich kleinere Schwankungen, die man unter der Bezeichnung *Nutation* zusammenfaßt; diese hat im größten Glied eine Periode von 18,5 Jahren.

Bei der *Chandler-Periode* von 12 bis 14 Monaten bewegt sich der Rotationspol auf einer komplizierten kreisähnlichen Bahn mit dem mittleren Radius von $\approx$ 10 m um einen erdfesten mittleren Pol. Die Ursache für diese Erscheinung liegt darin, daß Rotations- und Trägheitsachse der Erde nicht zusammenfallen und die Trägheitsachse durch Massenverlagerungen im Erdkörper und in der Atmosphäre zeitlichen Änderungen unterworfen ist.

Aus der Analyse des Verlaufes der Polbewegung versucht man, Rückschlüsse auf die elastischen Parameter des Erdkörpers (Nachgiebigkeit, *Righeit*) und auf geodynamische Prozesse im Erdinneren zu ziehen.

Neben diesen periodischen Schwankungen verringert sich die Rotationsgeschwindigkeit kontinuierlich durch die *Gezeitenreibung*.

10.2.1.3. Das Geoid. Eine Fläche, die überall senkrecht zur jeweiligen Lotrichtung verläuft, heißt *Niveaufläche der Schwerkraft*. Eine solche Fläche, die mit der ungestörten Oberfläche der Ozeane zusammenfällt und unter den Festlandsmassen etwa durch ein enges System von Kanälen weitergeführt zu denken ist, wird seit 1972 als *Geoid* (ideale Erdgestalt) bezeichnet. Die Dimensionen des Geoids werden für geodätische Höhenbestimmungen benötigt. Damit ergibt sich über das rein Geometrische hinaus ein wichtiger physikalischer Zusammenhang mit der Schwerkraft. Berechnet wird das Geoid aus Schweremessungen und Beobachtungen der Lotabweichungen auf der Erdoberfläche sowie aus der Analyse der Bahnelemente erdnaher künstlicher Satelliten.

Die Abweichungen des Geoids vom idealen Rotationsellipsoid heißen *Geoidundulationen*. Sie haben ihre Ursache in der inhomogenen Massenverteilung im Erdmantel. Die Extremwerte liegen bei $\mp$ 100 m. Diese relativ geringen Abweichungen werden als Anzeichen dafür gewertet, daß das isostatische Prinzip bei der Entwicklung der Erdfigur eine wichtige Rolle spielte; gegenüber langzeitigen Beanspruchungen verhält sich die Masse des Erdinneren wie ein plastischer Körper.

Bildet man die Mittelwerte der Geoidundulationen über alle Längengrade, so ergibt sich das Bild einer „birnenförmigen" Erde. Ausbuchtungen um $\approx$ 18 m am Nordpol und Einbuchtungen am Südpol um $\approx$ 26 m.

10.2.2. Erdmagnetismus

Die Erde ist ein schwacher Magnet. Die magnetische *Totalintensität* liegt zwischen 30 µT (Mikrotesla) am magnetischen Äquator und 69 µT in der Nähe der Magnetpole. Zur Beschreibung der Vektorgröße des *Magnetfeldes* ist die Angabe von drei Komponenten erforderlich. Die meistgebrauchten Komponenten, die Elemente des Feldes, sind die Beträge der *Total-*, der *Vertikal-* und der *Horizontalintensität* sowie die *Deklination* und die *Inklination* (Abb. 10.2.2–1). Die Deklination oder *Mißweisung* ist die Abweichung des magnetischen Meridians vom geografischen Meridian. Sie wird im Uhrzeigersinn positiv gerechnet.

10.2.2.1. Das geomagnetische Dipolfeld. Das auf der Erdoberfläche beobachtete *Magnetfeld* hat große Ähnlichkeit mit dem Feld auf der Oberfläche einer nicht magnetisierbaren Kugel, in deren Mittelpunkt man sich einen Stabmagneten, einen magnetischen Dipol, angeordnet denkt. Das geomagnetische *Hauptfeld* ($\approx$ 94 % des gesamten Feldes) wird deshalb auch *Dipolfeld* genannt. Es ändert sich zeitlich nur sehr langsam. Als *Säkularvariation* bezeichnet

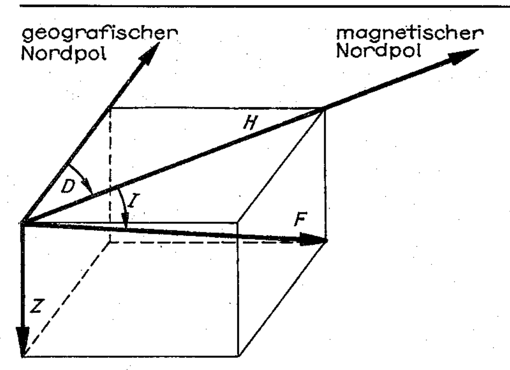


Abb. 10.2.2–1 Elemente des erdmagnetischen Feldes; *F* Totalintensität, *H* Horizontalintensität, *Z* Vertikalintensität, *D* Deklination, *I* Inklination

man die über längere Zeiträume ($\geqq 10^3$ Jahre) gleichgerichtete Veränderlichkeit.

Die Achse des Dipols ist gegen die Rotationsachse der Erde um 11,6° geneigt. Sein Südpol zeigt auf einen Punkt in 78,4° n. Br. und 69,0° w. L. Auf diese Achse bezieht man die geomagnetischen Koordinaten. Das *magnetische Moment* beträgt $\approx 8{,}0 \cdot 10^{15}$ Wbm. Die Feldkomponenten ergeben sich durch Differentiation aus der magnetischen Potentialfunktion $\varphi_p = \boldsymbol{M} \cdot \boldsymbol{r}/r^3$. Diese Formel kann man benutzen, um in erster Näherung die Elemente des Feldes im Punkte *P* der Erdoberfläche zu berechnen, z. B. die Vertikalintensität $Z_p = \partial\varphi_r/\partial r = 2M \sin\varphi/r^3$ (Abb. 10.2.2–2).

Das geomagnetische Dipolfeld und die *Säkularvariation* haben ihre wahrscheinliche Ursache an der Grenze Erdmantel – äußerer Erdkern, in Tiefen zwischen 2 900 und 5 100 km. Überlagert wird das Dipolfeld vom räumlich stark schwankenden Krustenfeld und von den Wirkungen des *magnetischen Außenfeldes* mit seinen zeitlichen Variationen.

10.2.2.2. Das magnetische Außenfeld. Das *magnetische Außenfeld* entsteht durch elektrische Ströme in den ionisierten Bereichen der Hochatmosphäre und in der Magnetosphäre. Seine zeitlichen Änderungen sind hauptsächlich die periodischen Variationen und die geomagnetischen Pulsationen. Unregelmäßige Variationen großer Intensität werden als *magnetische Gewitter* oder *Stürme* bezeichnet. Sie treten in Verbindung mit *Polarlichtern* auf und sind eine Folgeerscheinung von solaren Strahlungsausbrüchen (Elektronen und Protonen). Die Intensitäten erreichen an der Erdoberfläche kurzzeitig 1 µT. Die *Magnetosphäre* der Erde resultiert aus der Wechselwirkung des erdmagnetischen Feldes mit heißem Magnetoplasma, das als *Sonnenwind* ständig auf die Erde zuströmt. Die Leistung dieses Sonnenwindes beträgt durchschnittlich $2 \cdot 10^{10}$ kW. Dadurch wird das einfache Dipolfeld der Erde auf drastische Weise deformiert. Auf der Sonnenseite wird die Magnetosphäre auf ≈ neun Erdradien zusammengedrückt, auf der Nachtseite der Erde zieht der Sonnenwind das Erdmagnetfeld zu einem riesigen Schweif bis zu einer Entfernung von ≈ 1 000 Erdradien auseinander (vgl. Abb. 10.1.4–5).

10.2.2.3. Paläomagnetismus. Das erdmagnetische Feld hat im Laufe der Erdgeschichte mehrfach seine Richtung gewechselt. Relativ ruhige Epochen werden durch kurzperiodische Umkehrungen unterbrochen. Der letzte *Polaritätswechsel* hat vor ≈ 12 000 Jahren stattgefunden. Neben diesen Änderungen gibt es noch eine säkulare *Polwanderung*: So lag der geomagnetische Südpol vor ≈ 600 Millionen Jahren nordöstlich von Australien.

Die paläomagnetischen Untersuchungen stützten sich auf den „fossilen" Gesteinsmagnetismus. Gesteine mit remanent magnetisierbaren Mineralen erhalten bei ihrer Entstehung unter der Einwirkung des geomagnetischen Feldes eine natürliche *remanente Magnetisierung*, die im Labor gemessen werden kann. Beispielsweise bildet sich eine stark thermoremanente Magnetisierung aus, wenn magmatische Gesteine bei ihrer Abkühlung die Néel- bzw. Curie-Temperatur durchlaufen. Diese Temperatur hängt in hohem Maße von der stofflichen Zusammensetzung ab; sie liegt bei vielen Titanomagnetiten zwischen 380 K und 630 K, beträgt ≈ 600 K für Magnetkies, 750 K für Magnetit und 950 K für Hämatit-Modifikationen.

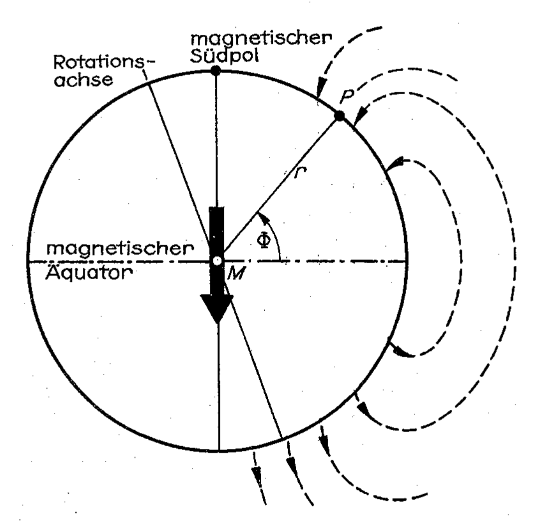


Abb. 10.2.2–2 Das magnetische Dipolfeld der Erde

10.2.3. Seismologie

Die *Seismologie*, die (*Erdbebenkunde*) untersucht Ursachen, Wirkungen und geografische Verbreitungen natürlicher Erdbeben. Aufgrund der unterschiedlichen Ausbreitungsformen elastischer Wellen liefert sie wichtige Informationen über den Schalenbau und die stoffliche Beschaffenheit des Erdkörpers. Im Gegensatz zur Seismologie bedient sich die *Seismik* als Teilgebiet der angewandten Geophysik künstlicher Energiequellen, wie Sprengungen, Fallgewicht oder Schwingungsmaschinen. Die Seismik wird zur Erforschung des geologischen Aufbaues der Erdkruste und besonders zur Erkundung von Erdöl- und Erdgaslagerstätten eingesetzt. Eine spezielle Aufgabe der sogenannten „forensischen Seismologie" ist die Aufdeckung und Ortung unterirdischer Kernexplosionen.

10.2.3.1. Ausbreitung elastischer Wellen im Erdkörper. Von einem *Erdbebenherd*, dem *Hypozentrum*, gehen verschiedene elastische Wellen nach allen Richtungen aus. Auf gekrümmten Bahnen durchlaufen die *Longitudinal-* oder *Kompressionswellen* und die *Transversal-* oder *Scherwellen* als Raumwellen das Erdinnere; an der Erdoberfläche breiten sich mit einer geringen Eindringtiefe die *Oberflä-*

chenwellen, *Love-* und *Rayleigh-Wellen*, aus. Alle diese Wellenarten erreichen eine seismologische Station zu verschiedenen Zeiten. Die Longitudinalwellen haben die größte Geschwindigkeit; nur etwa halb so groß ist die der Rayleigh-Wellen. An Grenzflächen in der Tiefe, wie auch an der Erdoberfläche selbst, werden die seismischen Wellen gebrochen, reflektiert und teilweise in eine andere Wellenart transformiert (Entstehung von Wechselwellen). Eine solche Diskontinuität erster Ordnung ist in 2 900 km Tiefe die Kern-Mantel-Grenze, an der sich Ausbreitungsgeschwindigkeit, Dichte und elastische Eigenschaften sprunghaft ändern (vgl. 10.2.5.2. und 10.1.4.1.).

Die schnellen Longitudinalwellen werden, weil sie als erste Einsätze im Seismogramm erscheinen, gewöhnlich *P-Wellen* („undae primae", lat., erste Welle) genannt. Entsprechend heißen die Transversalwellen kurz auch S-Wellen („undae secundae"). Für den kürzesten Weg durch den Erdmittelpunkt zum Antipoden benötigt die P-Welle eine Laufzeit von ≈ 20 min.

10.2.3.2. Seismogrammauswertung. Die Registrierung seismischer Wellen erfolgt je nach Gerätetyp auf mechanischem, optischem oder elektromagnetischem Weg mit Verstärkungen bis auf das Millionenfache. Das meßbare Spektrum umfaßt den Bereich von 10 bis 0,001 Hz. Um die Bodenbewegung möglichst getreu aufzeichnen zu können, wird ihr Vektor in drei translatorische Komponenten zerlegt, so daß ein Vertikal- und zwei Horizontalseismografen (für N-S- und O-W-Richtung) einen kompletten Gerätesatz bilden. Zur Analyse des gesamten aufgezeichneten Wellenzuges wird in zunehmendem Maße die moderne Rechentechnik mit digitaler Datenerfassung eingesetzt.

Seismogramme enthalten Informationen über die Lage des Bebenherdes, den Bebenmechanismus und die physikalischen und geometrischen Eigenschaften des Wellenweges. Zur Standardauswertung gehört die Bestimmung der Zeitpunkte der Welleneinsätze sowie der Amplituden und Perioden der einzelnen Wellengruppen. Je größer die Entfernungen vom Herd sind, um so deutlicher und weiter auseinandergezogen werden die einzelnen Einsätze bzw. Einsatzgruppen wegen der unterschiedlichen Wellengeschwindigkeiten; dabei wachsen die Perioden der Wellen von Sekunden- auf Minutenlänge an.

Aus den gleichzeitigen Registrierungen an mehreren seismologischen Stationen werden die Koordinaten des Bebenherdes berechnet und die von der epizentralen Entfernung abhängigen Laufzeiten diskreter Wellengruppen als Laufzeitkurve dargestellt. Die Laufzeitkurven bilden die Grundlage für die Rekonstruktion der einzelnen Wellenwege.

Die maximale Amplitude A einer seismischen Welle gegebener Perioden T liefert ein quantitatives Maß für die von einem Beben freigesetzte Energie, $E \sim (A/T)^2$. Für Klassifikationen ist der Logarithmus der Energie eine geeignete Größe:

$\lg E = a' + b' \lg (A/T) = a + bM$. In diesem Ausdruck bedeutet M die von Ch. F. Richter (1935) eingeführte *Magnitude*. In den Koeffizienten stecken Standardkorrekturen, die die Epizentralentfernung, die Herdtiefe und die verwendete Wellenart berücksichtigen. Die Magnitude kann aus Aufzeichnungen von Raum- oder Oberflächenwellen berechnet werden. Beispielsweise gilt für Oberflächenwellen von Flachbeben die Beziehung $\lg E = 11{,}8 + 1{,}5\, M_S$ (S = surface, engl., Oberfläche); E in 10^7 Joule.

10.2.3.3. Erdbeben. Über 90 % aller *Erdbeben* sind *tektonischen* Ursprungs, entstehen bei Gebirgsbildungsprozessen, bei Lage- und Gestaltsänderungen in der Lithosphäre.

Durch Materieströmungen im Erdmantel wird die darüberliegende feste Gesteinshülle der Erde auf Zug, Druck und Scherung beansprucht. In der Lithosphäre entwickelt sich so ein mechanischer Spannungszustand. Örtlich und besonders an den Rändern sich bewegende Geoplatten kann der Schwellenwert der Bruchfestigkeit erreicht werden; dort kommt es zum Bruch des Erdkrustenstückes, wobei die angehäufte Deformationsenergie plötzlich freigesetzt wird und ein Spannungsausgleich eintritt.

Eine hinsichtlich Häufigkeit und Intensität untergeordnete Rolle spielen die *vulkanischen Beben*, die bei explosiven Vulkanausbrüchen auftreten, und noch weniger häufig sind die *Einsturzbeben*, die auf den Zusammenbruch unterirdischer Hohlräume zurückzuführen sind.

Zur Einschätzung der Erdbebenstärke und zur Schnellinformation über das Ausmaß von Schäden sind Intensitätsskalen entwickelt worden, die von sichtbaren und fühlbaren Wirkungen im Epizentralgebiet ausgehen und häufig auch die Situation und die Belange des jeweiligen Landes berücksichtigen. Weltweit im Gebrauch ist die zwölfstufige modifizierte *Mercalli-Skala* (MMS). (Benannt nach dem Seismologen G. Mercalli, 1850–1914.) Daneben wird immer häufiger, und für geophysikalische und bautechnische Berechnungen fast ausschließlich, die aus instrumentellen Aufzeichnungen abgeleitete quantitative Magnitude (vgl. 10.2.3.2.) verwendet. Die auch als *Richter-Skala* bezeichnete Magnitudenverteilung läßt sich mit Intensitätsskalen in Beziehung bringen. Magnitudenwerte über 9,0 wurden bisher in historischer Zeit von keinem Erdbeben erreicht.

10.2.4. Geothermie

10.2.4.1. Der Wärmefluß zur Erdoberfläche. Die Temperaturstruktur ist die am unsichersten bekannte physikalische Zustandsgröße des Erdkörpers, weil sie aus anderen besser bekannten (Druck, Dichte) wegen des nur losen Zusammenhanges mit diesen nicht erschlossen werden kann. Direkt an oder unter der Erdoberfläche wird der Erdwärmestrom durch Tages- und Jahresgang des Wärmestromes aus der Sonnenstrahlung überlagert, der unter 20 m Tiefe aber praktisch verschwindet.

Der *Wärmefluß* $Q = -\lambda \operatorname{grad} T$ (λ = Wärmeleitfähigkeit; für Gesteine 10 bis $65 \cdot 10^{-3}$ J/cm s K) beträgt im Mittel über die gesamte Erde $\bar{Q} = (6{,}3 \mp 0{,}45)\, 10^{-6}$ J [cm² s ≙ Wcm⁻²]. Damit fließen, obwohl dieser Wärmestrom vier Größenordnungen unter der Solarkonstanten (vgl. 10.4.1.) liegt, jährlich 10^{21} J aus dem Erdkörper heraus. Nach Beobachtungen in Bergwerken und Bohrlöchern beträgt die mittlere Temperaturzunahme mit der Tiefe 1 K/30 m; dieser Wert bis zum Erdkern angesetzt würde zu unrealistisch hohen Temperaturen führen. Als Grenzwert der Temperaturen im Erdinneren kann die Schmelztemperatur angesetzt werden. Aus der Clausius-Clapeyronschen Gleichung erhält man in Verbindung mit der statischen Grundgleichung und mit charakteristischen Werten für die Silikate (Schmelztemperatur $\approx 1\,570$ K, Schmelzwärme ≈ 420 J/g) eine Schmelztemperaturabhängigkeit von der Tiefe von ≈ -3 K/km. Zur Modellierung der *Temperaturstruktur des Erdinneren* nimmt man an, daß vertikaler Temperatur- und Schmelztemperaturverlauf sich im Bereich des (plastischen) oberen Erdmantels nähern (Abb. 10.2.4–1), sich an der Erdkerngrenze treffen und daß im äußeren Kern die Temperatur über die Schmelztemperatur ansteigt. Im Kern wird eine dem adiabatischen Temperaturverlauf entsprechende thermische

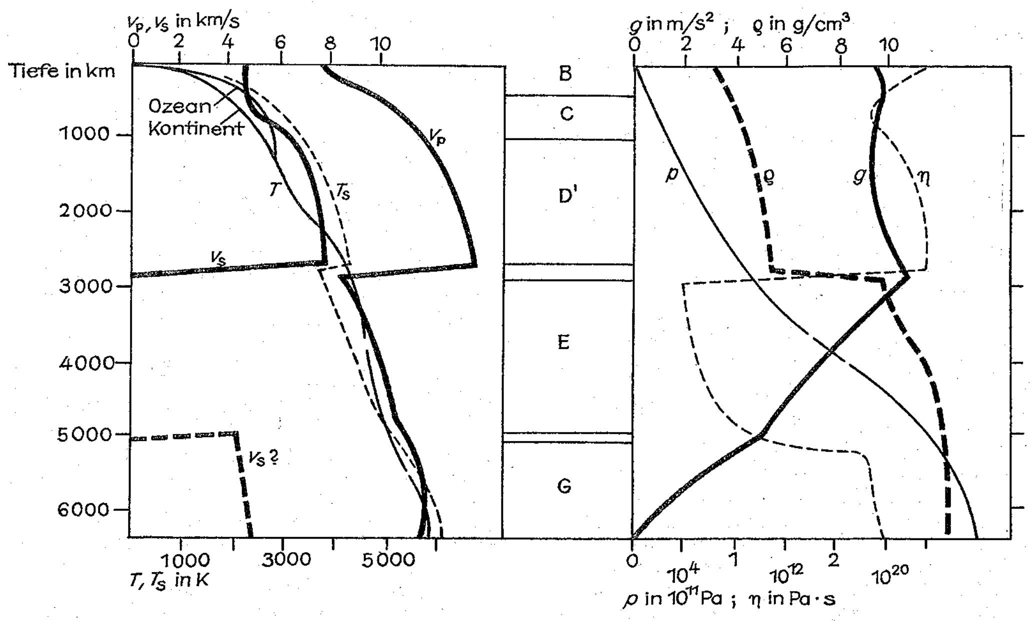


Abb. 10.2.4–1 Tiefenabhängigkeiten physikalischer Zustände im Erdkörper: v_p, v_s = Geschwindigkeiten der P- und S-Wellen, T, T_s = Temperatur und Schmelztemperatur, g = Schwerebeschleunigung, ϱ = Dichte, p = Druck, η = dynamische Viskosität, B ... G = Bullen-Zonen

Struktur angenommen, deren Wert sich (ausgehend vom 1. Hauptsatz, der Thomson-Gleichung und der statischen Grundgleichung) zu $\mathrm{d}T/\mathrm{d}z_a = -(g\alpha/c_p)\,T$ ergibt und mit den Werten des äußeren Erdkerns ($g = 0{,}8\cdot 10^3\ \mathrm{cm/s^2}$, $\alpha = 10^{-5}\ \mathrm{K^{-1}}$, $c_p = 0{,}8\ \mathrm{J/g\cdot K}$, $T = 5\,000\ \mathrm{K}$) $\mathrm{d}T/\mathrm{d}z_a = -0{,}5\ \mathrm{K/km}$ beträgt.

Der wichtigste Prozeß für den *Wärmetransport* ist die Wärmeleitung. Daneben wirkt noch die Strahlung und in Gebieten, die durch Vulkanismus und Tiefenwässer gestört sind, kommen noch konvektive Wärmeströme hinzu.

10.2.4.2. Die Wärmeproduktion im Erdkörper. Wie Abschätzungen des Wärmeflusses zeigen, kann dieser höchstens zu 20 % aus der „Ursprungswärme" stammen, die mit der Planetenbildung akkumulierte. Der größte Teil stammt aus der Wärmeproduktion durch radioaktiven Zerfall von Gesteinen. Als Wärmeproduzenten wirken Uranium, Thorium und das Kaliumisotop ^{40}K. Die spezifischen freigesetzten Wärmeleitungen betragen für natürliches Uranium (99,28 % ^{238}U, 0,71 % ^{235}U, 0,01 % ^{234}U) $9{,}38\cdot 10^{-8}$ W/g, für Thorium (^{232}Th) $2{,}64\cdot 10^{-8}$ W/g und für natürliches Kalium (0,011 % ^{40}K) $3{,}56\cdot 10^{-12}$ W/g. Hieraus lassen sich die in Tab. 10.2.4–2 zusammengestellten spezifischen Wärmeproduktionen von Gesteinen berechnen; 1 km³ Granit liefert danach 2 770 W.

Setzt man für die gesamte Erde chondritisches Material an, ergibt sich für die Erdoberfläche ein *Wärmefluß* $Q = (M_E\cdot q)/4\pi R^2 = 6{,}1\cdot 10^{-6}\ \mathrm{W/cm^2}$ in guter Übereinstimmung mit dem beobachteten Mittelwert. Aber chondritisches Material existiert erst vom oberen Mantel an aufwärts. Doch auch eine 15 km dicke Granitschicht produziert bereits den beobachteten Wert. Der an der Erdoberfläche existierende Wärmefluß ist somit aus radioaktiver Produktion in Kruste und oberem Mantel voll erklärbar.

10.2.5. Aufbau des festen Erdkörpers

10.2.5.1. Modellvorstellung vom Schalenaufbau. Bei der ersten Auswertung von Seismogrammen und der Aufstellung von Laufzeitkurven zu Beginn des 20. Jahrhunderts er-

Tab. 10.2.4–2 Spezifische Wärmeproduktion q wichtiger Gesteine durch radioaktiven Zerfall (q in 10^{-14} W/g, Konzentration in g Substanz/t Gestein, ϱ = Dichte in g/cm³)

Gestein	ϱ	Uranium Konzentration	q	Thorium Konzentration	q	Kalium Konzentration	q	gesamt q
Granit	2,6	4,75	44,38	18,5	48,78	37 900	13,48	106,64
Krustengestein	2,85	2,0	18,76	7,5	20,18	18 000	6,41	45,35
Basalt	3,3	0,6	5,61	2,7	7,12	8 400	2,97	15,70
Dunit	3,9	0,001	0,008	–	–	10	0,004	0,013
Chondrite	3,2	0,012	0,113	0,04	0,105	845	0,301	0,519

kannte man bereits, daß die Geschwindigkeitsverteilung seismischer Wellen innerhalb des Erdkörpers nicht gleichmäßig ist. In verschiedenen Tiefenbereichen gibt es Diskontinuitäten, an denen die Geschwindigkeit sich abrupt ändert, die einfallenden elastischen Wellen reflektiert und refraktiert werden und zum Teil auch in andere Wellenarten wechseln. Diese Beobachtung führte zur Erkenntnis vom schalenförmigen Aufbau der Erde mit den drei Hauptbereichen *Erdkruste*, *Erdmantel* und *Erdkern*. Eine Untergliederung stammt von dem neuseeländischen Seismologen K. E. Bullen (1963); die nach ihm benannten Zonen zeichnen sich durch eine charakteristische Geschwindigkeitsverteilung aus und bilden den Gegenstand spezieller Untersuchungen der Physik und der Chemie des Erdinneren (vgl. Abb. 10.1.4–1 und 10.2.4–1).

10.2.5.2. Physikalische Zustände als Funktion der Tiefe. Die *rheologischen Parameter* (Parameter des Fließverhaltens) der Materialien des Erdinneren können nur abgeschätzt werden. Hinweise liefern Rotationsschwankungen der Erde, Gezeiteneffekte, Eigenschwingungen des Erdkörpers und die Dämpfung seismischer Wellen. Danach zeigt z. B. *dynamische Viskosität* im oberen Mantel mit 10^{19} bis 10^{20} Pa · s ein relatives Minimum (vgl. Abb. 10.2.4–1), eine Feststellung, die für die Beurteilung geotektonischer Prozesse in der Lithosphäre von großer Bedeutung ist. Die Dichte im tieferen Erdinneren berechnet man aus den Angaben über die Dichte der Krustengesteine und über die mittlere Dichte der gesamten Erde unter Berücksichtigung der Ausbreitungsgeschwindigkeit seismischer Wellen. Druck und Schwerebeschleunigung lassen sich dann aus der Dichteverteilung bestimmen. Über den Temperaturverlauf und die elektrische Leitfähigkeit im Erdinneren kann man gegenwärtig nur Wahrscheinlichkeitsaussagen machen. So liegt die Größenordnung für die Temperatur und für den spezifischen elektrischen Widerstand im Erdkern bei $\approx 5\,000$ K bzw. $10^{-6}\,\Omega$m.

Über die *stoffliche Zusammensetzung* des tieferen Erdinneren bestehen viele unterschiedliche Vorstellungen, da es zwischen den physikalischen Zustandsgrößen und den stofflichen Zuständen der Materie keine eindeutigen Beziehungen gibt. Innerhalb des Erdmantels treten wahrscheinlich weniger chemische als vielmehr Phasenänderungen auf. Der untere Erdmantel dürfte vorwiegend aus Silikaten mit einem größeren Anteil an Eisenoxiden zusammengesetzt sein. Bei dem stark ansteigenden Druck rücken die atomaren Bausteine im Kristallgitter näher zusammen, werden dichter gelagert und bilden sogenannte *Hochdruckphasen*. Hinsichtlich des Erdkerns fordern die meisten geochemischen Hypothesen eine eisenreiche Zusammensetzung in metallischer Phase. Für den äußeren Erdkern muß jedoch ein metallisch-quasiflüssiger Zustand angenommen werden, da er Transversalwellen so gut wie nicht hindurchläßt und als Hauptquelle des geomagnetischen Innenfeldes (nach dem Dynamoprinzip) angesehen wird.

10.2.6. Geophysikalische Erkundung mineralischer Ressourcen

Wissenschaft und Praxis der Erkundung von nutzbaren Lagerstätten in der Erdkruste mit Hilfe physikalischer Meßmethoden sind Arbeitsgebiete der *angewandten Geophysik*. Die Anwendung beruht auf den unterschiedlichen physikalischen Eigenschaften der verschiedenen Gesteinsarten und der Bodenschätze, die meßbare Effekte an der Erdoberfläche verursachen. Viele Lagerstätten sind geophysikalisch nicht direkt nachweisbar. Entweder sind die wirksamen Massen zu gering oder die Vorkommen liegen zu tief. Das trifft vor allem für die Erdöl-, Erdgas-, Kohle- und für solche Erzlagerstätten zu, bei denen die interessierenden Minerale in zu winziger Verteilung im Gestein vorliegen. Stets aber lassen sich Aussagen über die begleitenden geologischen Verhältnisse machen, an welche ein bestimmtes Vorkommen gebunden ist, z. B. über Lage und Begrenzung einzelner Schichten, über Verwerfungen, Aufwölbungen u. a. Die geophysikalischen Erkundungsergebnisse werden häufig dazu verwendet, die günstigsten Ansatzpunkte für geologische Aufschlußbohrungen herauszufinden.

Die geophysikalischen Erkundungsmethoden lassen sich in zwei Gruppen einteilen: Die erste Gruppe bedient sich *spontaner Kraftfelder*, wie Schwere, Magnetfeld, Wärmefluß, Radioaktivität oder elektrische Felder. Ausgenutzt wird die Fernwirkung, die Gesteine und Einlagerungen im Untergrund ausüben. Bei der zweiten Gruppe der *provozierten Felder* werden künstliche Bodenerschütterungen oder elektrische Felder in den zu untersuchenden Gebieten erzeugt und die Reaktion des Untergrundes auf diese „Befragung" gemessen.

Die *seismische Methode* besteht in der Beobachtung der Ausbreitungsvorgänge künstlich, meist durch Sprengungen, erzeugter Erdbebenwellen. Bei der Untersuchung geringer Tiefen, besonders des Baugrundes, werden zur Anregung der Wellen auch Hammerschläge oder Fallgewichte verwendet. Zur sprengstofflosen Energieanregung werden Schwingungsmaschinen, Vibratoren mit regelbarer Frequenz, zunehmend auch für die Erkundung sehr großer Tiefen (einige 10 km) eingesetzt, seitdem eine digitale Registrierung und Signalbearbeitung möglich geworden ist.

Für die Ausbreitung elastischer Wellen gelten ähnliche Gesetze wie in der Optik. An geologischen Schichtgrenzen, an denen sich Fortpflanzungsgeschwindigkeit und Dichte ändern, treten Refraktionen und Reflexionen auf, und ein Teil der ausgestrahlten Energie gelangt wieder an die Erdoberfläche. Je nach der Aufstellungsweite der Empfangsgeräte, der *Geophone*, und der speziellen Registriertechnik unterscheidet man Refraktions- und Reflexionsmethode. Durch Auswertung der Laufzeiten seismischer Wellen erhält man nach der *Refraktionsseismik* einen Überblick über die Geschwindigkeitsverteilung im Untergrund des Meßgebietes. Aufragungen oder Einlagerungen von Gesteinen mit hoher Schallgeschwindigkeit werden durch relativ kurze Laufzeiten erkannt. Aufgabe der *Reflexionsseismik* ist die Bestimmung der Lage einzelner seismischer Grenzflächen in annähernd horizontal geschichteten Medien aus der Beobachtung reflektierter Wellen. Mit diesem Verfahren können Tiefenbestimmungen für die Reflexionshorizonte sehr genau ausgeführt und der Schichtenverlauf im Untergrund abgebildet werden.

Die *Seismik* liefert von allen geophysikalischen Verfahren die genauesten Informationen über die Lagerungsform und die Mächtigkeit von Sedimentgesteinen. Sie wird besonders zur Erkundung sedimentärer Lagerstätten von Erdöl, Erdgas, Kohle und Salz eingesetzt.

Die Anwendung der *Gravimetrie*, Schweremessung, beruht darauf, daß die unterschiedliche Massenverteilung in der Erdkruste meßbare *Schwereanomalien*, Abweichungen vom normalen Schwerefeld, verursacht. Die Gravitationswirkung eines geologischen Körpers hängt von seiner Gestalt, der Lage und Tiefe, sowie von seinem Dichtekontrast zur Umgebung ab. Aus Größe und Form einer Schwereanomalie

läßt sich ableiten, ob im betreffenden Gebiet ein Massenüberschuß oder ein Massendefizit vorliegt und in welcher Tiefe die Massenunregelmäßigkeit liegt. Steilstehende Diskontinuitäten werden im Anomalienbild durch große Gradienten der Schwerestörung erkannt.

Für die Erkundung lokaler geologischer Strukturen (Verwerfungen, Antiklinalen, Sättel, Horste u. a.) sowie von Salzstöcken und ausgedehnten Erzlagerstätten wird eine Meßgenauigkeit von 10^{-5} cm/s^2 für die relative Schwereänderung angestrebt. Die Messungen erfolgen mit dem *Gravimeter*, einer hochempfindlichen Federwaage, bei der man die Längenänderung einer elastischen Feder unter dem Einfluß einer Schwereänderung beobachtet.

Klassisches Einsatzgebiet der *Geomagnetik* ist die Erkundung von Eisenerzlagerstätten. Die heute eingesetzten Meßinstrumente arbeiten auf elektronischer Basis, z. B. die Atom- oder Kernresonanzmagnetometer, und erreichen eine Genauigkeit von 0,1 nT (Nanotesla). Mit ihrer Hilfe kann die geomagnetische Vermessung auch vom Flugzeug oder vom Satelliten aus erfolgen.

Die Abweichungen vom großregionalen Feldverlauf, die geomagnetischen Anomalien, werden durch magnetisch stärker wirksame Gesteine und Erze verursacht. Träger der induzierten und remanenten Magnetisierung sind in erster Linie die Minerale Magnetit, Titanomagnetit und Magnetkies. Die Tiefenreichweite der geomagnetischen Methode ist auf ≈ 15 bis 20 km beschränkt, weil in dieser Tiefe die Néel-Temperatur (≈ 850 K, Verschwinden des Ferrimagnetismus) erreicht wird.

Die *Geoelektrik* als Prospektionsverfahren dient hauptsächlich dazu, Erz-, Wasser- und Werksteinvorkommen zu erkunden. Ihre Anwendung beruht auf den unterschiedlichen elektrischen Eigenschaften geologischer Körper. Die zahlreichen Spezialverfahren lassen sich in zwei Gruppen einteilen. Bei der einen Gruppe werden natürliche elektrische Erdströme oder die z. B. an sulfidischen Erzkörpern auftretenden Eigenpotentiale gemessen. Größere Bedeutung hat die andere Gruppe, bei der man die Ausbreitung künstlich geschaffener elektrischer Gleich- oder Wechselstromfelder untersucht. Dem Erdboden wird der Strom entweder galvanisch (mit Elektroden) oder, besonders bei hochohmiger Beschaffenheit des Bodens, induktiv (mit Induktionsrahmen) zugeführt. In unwegsamen Gelände kann die Lagerstättensuche auch vom Flugzeug aus erfolgen. Hierbei werden Sende- und Empfangsrahmen in konstantem Abstand über den Erdboden geführt und die Effekte von Wirbelströmen registriert, die sich in gutleitenden Erzkörpern ausbilden.

10.3. Physik der Hydrosphäre

Die Beschreibung der physikalischen Prozesse in der *Hydrosphäre* beginnt bei den physikalischen Eigenschaften des reinen Wassers und den Phasenumwandlungen. Zu diesen Eigenschaften zählen vor allem:

- Eis- und Siedepunkt ermöglichen unter irdischen Bedingungen neben flüssigem Wasser auch die Eis- und die Dampfphase,
- hohe spezifische Wärme (4,187 $J^{-1}K^{-1}$), Schmelzwärme (330,77 J/g) und Verdampfungswärme 2 512,2 J/g) verursachen eine hohe Wärmespeicherfähigkeit des Wassers und dämpfen Temperaturextrema,
- der Aufbau des Wassermoleküls (kleines Molvolumen, hohe Dielektrizitätskonstante) bewirken eine große chemische Aktivität (hohe Dissoziationskraft und hohes Lösungsvermögen),
- temperaturabhängige Relativanteile von Wassermolekülen und verschiedenen tetraedralen Strukturen führen zu maximaler Wasserdichte bei +4 °C und zu sprunghafter Volumenerhöhung um ≈ 9 % beim Gefrieren; damit werden schützende Gewässereisdecken gegen weitere starke Abkühlung gebildet.

10.3.1. Physik des Meeres

10.3.1.1. Eigenschaften des Meerwassers. Chemische Eigenschaften. Meerwasser besteht aus einer Lösung von Salzen, organischen Stoffen und Gasen und unlöslichen Suspensionen. Bei den gelösten anorganischen Stoffen werden Hauptbestandteile und Spurenstoffe unterschieden, wobei die Hauptbestandteile in sehr konstanter relativer Zusammensetzung enthalten sind (Tab. 10.3.1–1). Der mittlere *Salzgehalt* der Weltmeere beträgt 35 ‰; die Spurenstoffe machen nur < 0,1 % dieses Salzgehaltes aus. Die Bestimmung des Salzgehaltes erfolgt in der Regel über Chlortitration oder über genaue Messungen der elektrischen Leitfähigkeit.

Von den im Wasser gelösten Gasen haben Sauerstoff und Kohlendioxid besondere Bedeutung. O_2 wird im Wasser fotosynthetisch von den Pflanzen produziert und durch Atmung von Tieren und Pflanzen wieder verbraucht. CO_2 unterliegt komplizierten Abhängigkeiten und das CO_2-System des Meeres ist wichtiger Teil des globalen CO_2-Stoffaustausches zwischen den Geosphären.

Physikalische Eigenschaften. Der statische Druck in einer bestimmten Wassertiefe entspricht dem Gewicht der Wassersäule über der Flächeneinheit: $\bar{p} = g \bar{\varrho} z$ [Nm^{-2}]. Die

Tab. 10.3.1–1 Hauptbestandteile des Meerwassers bei 0 °C und mittlerem Salzgehalt von 35 ‰

	g/kg	Millimol/kg
Kationen		
Natrium	10,752	467,56
Kalium	0,39	9,98
Magnesium	1,295	53,25
Calcium	0,416	10,38
Strontium	0,013	0,15
Anionen		
Chlor	19,345	545,59
Brom	0,066	0,83
Fluor	0,001 3	0,07
Sulfat	2,701	28,12
Bicarbonat	0,145	2,38
Borsäure	0,027	0,44
Salze (in kg/m^3)		
$NaCl$	28,014	
$MgCl_2$	3,812	
$MgSO_4$	1,752	
$CaSO_4$	1,283	
K_2SO_4	0,816	
$CaCO_3$	0,122	
KBr	0,101	
$SrSO_4$	0,028	

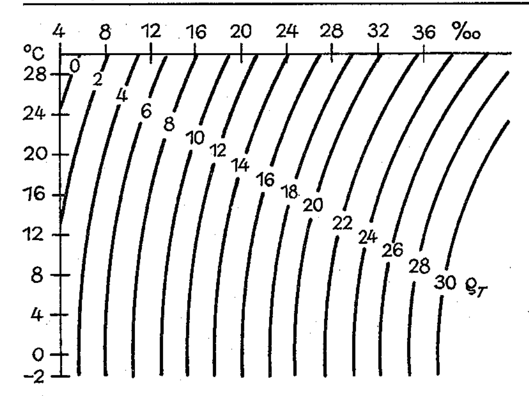


Abb. 10.3.1–2 Meerwasserdichte in Abhängigkeit von Salzgehalt und Temperatur (Angaben für $\delta_T = \varrho_{\text{Wasser}} - 10^3$ kg/m³)

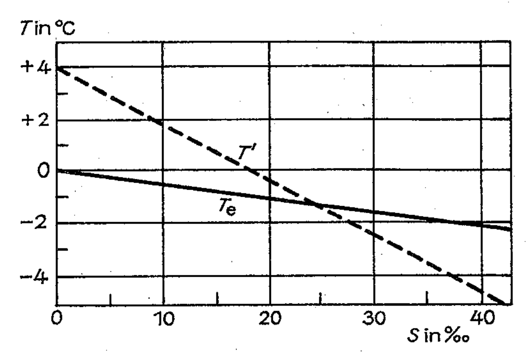


Abb. 10.3.1–3 Abhängigkeiten der Temperatur des Dichtemaximums T' und des Gefrierpunktes T_e vom Salzgehalt

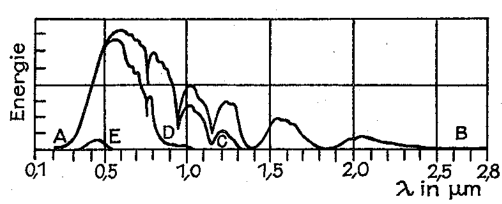


Abb. 10.3.1–4 Wellenlängenabhängige Eindringtiefen der Sonnenstrahlung; A–B Energieverteilung an der Meeresoberfläche, A–C in 1 cm, A–D in 1 m und A–E in 100 m Tiefe bei reinem Meerwasser

Dichte ϱ des Meerwassers hängt von der Temperatur, dem Salzgehalt und wegen geringer Kompressibilität des Meerwassers geringfügig auch vom Druck ab. Die Abb. 10.3.1–2 enthält *Wasserdichten* in Abhängigkeit von Temperatur und Salzgehalt (für $p = 0$) mit Angaben δ_T, wobei $\varrho(T, S) = 10^3 + \delta_T$ [kgm⁻³] gilt.

Mit der Zunahme des Salzgehaltes S verschieben sich die Temperaturen für das Dichtemaximum T' und für den Gefrierpunkt T_e gemäß Abb. 10.3.1–3. Solange der Gefrierpunkt T_e tiefer liegt als die Temperatur des Dichtemaximums T', das ist bis $S < 24{,}7$ ‰, wird bei Abkühlung von der Meeresoberfläche her zunächst vertikale Konvektion durch absinkendes kälteres Oberflächenwasser eintreten, bis sich nach Unterschreiten von T' Stabilität einstellt und sich bei Erreichen von T_e Oberflächeneis bildet. Bei größerem Salzgehalt wirkt sich dagegen wegen $T_e > T'$ Abkühlung der Oberfläche in fortgesetzter Umlagerung aus, und Eis kann sich erst dann bilden, wenn der gesamte Wasserkörper auf die Gefriertemperatur abgekühlt ist; das erklärt das stark verzögerte Gefrieren von Meerwasser.

Die auf die Meeresoberfläche treffende *Globalstrahlung* (direkte Sonnen- und diffuse Himmelsstrahlung) wird durch das Wasser in hohem Maße absorbiert und führt zu geringen, wellenlängenabhängigen Eindringtiefen (Abb. 10.3.1–4). Die *Albedo* (das Reflexionsvermögen) ist insgesamt relativ klein, unter 10 %, dabei u. a. abhängig von der Wellenlänge der Strahlung und vom Seegang. Dadurch entsteht eine hohe Umsetzungsrate von Strahlungs- in Wärmeenergie, und vertikale und horizontale *Wärmetransporte* mit den Wasserbewegungen sorgen weiter dafür, daß die Meere große Wärmespeicherfähigkeit haben und für eine Milderung örtlicher und zeitlicher Temperaturunterschiede sorgen.

Die spektrale Verteilung der optischen Durchlässigkeit der Wassermassen ist sehr unterschiedlich. Besonders Küstengewässer erreichen eine 30- bis 40fache *Extinktion* gegenüber reinem Wasser.

Schallwellen breiten sich im Wasser als Longitudinalwellen über weite Strecken aus und werden zur Informationsgewinnung und Nachrichtenübermittlung verwendet. Die Schallgeschwindigkeit c_s variiert in Abhängigkeit von Temperatur, Salzgehalt und Druck etwa zwischen 1 400 und 1 550 m/s. Sie ist bestimmbar nach der Beziehung $c = (c_p/c_v \varrho K)^{1/2}$, wobei K die druckabhängige adiabatische Kompressibilität darstellt (Abb. 10.3.1–5).

Die Schallausbreitung im Meer unterliegt nach den Gesetzen der Wellenausbreitung einer Absorption der Schallenergie, die mit dem Quadrat der Frequenz und proportional zur Viskosität wächst. Nach den Beobachtungen beträgt

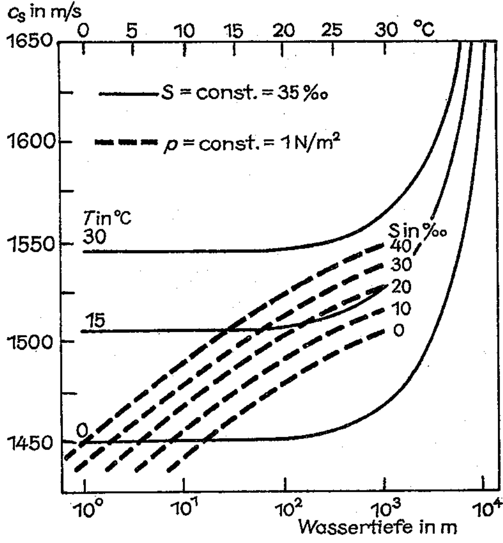


Abb. 10.3.1–5 Schallgeschwindigkeiten c_s im Meer in Abhängigkeit von Druck (Wassertiefe), Temperatur und Salzgehalt

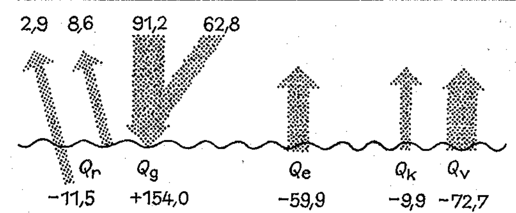


Abb. 10.3.1–6 Schema des mittleren Wärmehaushaltes der Meeresoberfläche (Zahlenangaben in W/m²)

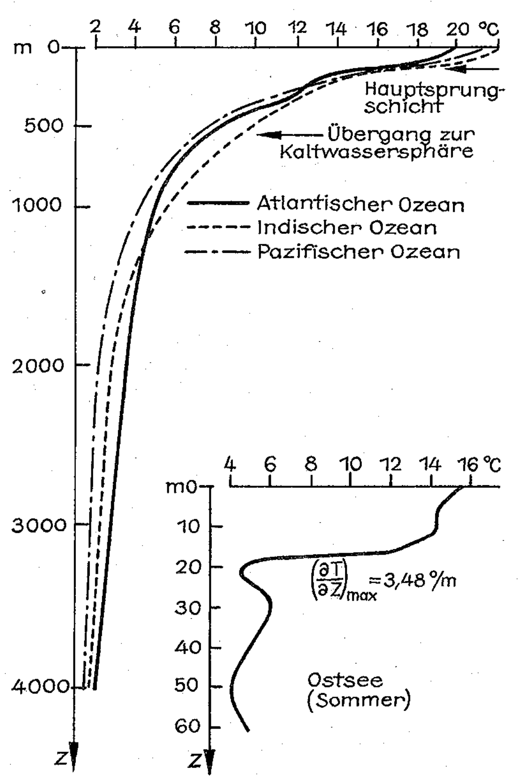


Abb. 10.3.1–7 Typische sommerliche Temperaturschichtung in der Ostsee und mittleres Temperaturprofil in den Ozeanen

der Dämpfungskoeffizient für Schallwellen des Bereiches $f = 16 \dots 60$ kHz $s = 0{,}036 f^{3/2}$ db/km. Refraktion, Reflexion und Streuung beeinflussen die Schallausbreitung außerdem, so daß wegen der Veränderlichkeiten von ϱ bzw. T und S räumlich und zeitlich sehr variable Schallausbreitungsformen auftreten. Beispielsweise bilden sich bei markanten thermohalinen Schichtstrukturen *Schallkanäle* mit zum Teil beträchtlicher Bündelung der Schallenergie aus. Insgesamt ist die Reichweite der Schallausbreitung im Meer etwa um das 10^3fache größer als in Luft.

10.3.1.2. Wärmehaushalt und Temperaturregime des Meeres. Die im Meer herrschenden Zustände und Prozeßabläufe sind Ausdruck des sehr verzweigten Systems der Energieumsetzungen. Hauptquelle ist die als Globalstrahlung dargebotene Sonnenenergie. Quellen kleinerer Größenordnungen bilden der Erdwärmestrom, biologische und chemische Umsetzungen, radioaktiver Zerfall von Stoffen im Meerwasser, Dissipation von Bewegungsenergie; diese sekundären Quellen zusammen erreichen höchstens den 10^{-3} bis 10^{-2}fachen Betrag der solaren Energiezufuhr.

Die mittlere Strahlungsbilanz der Meeresoberfläche (Abb. 10.3.1–6) ist positiv, d. h., die Energiezufuhr aus der Globalstrahlung Q_g ist größer als der effektive Energieverlust durch die langwellige Eigenstrahlung Q_e. Erst durch die *Wärmeströme*, durch den Transport fühlbarer Q_k und latenter, im Wasserdampf über die Verdunstungswärme gebundener Energie Q_v wird der Wärmehaushalt insgesamt ausgeglichen.

Der Wärmetransport im Meer erfolgt in entscheidender Weise über die vertikale Konvektion. Hierunter sind Austauschprozesse zu verstehen, die dann unter der Wirkung der Schwerkraft einsetzen, wenn spezifisch schwerere Wasserteilchen über leichteren liegen. Ursachen einer solchen Konvektion liegen in oberer Abkühlung oder Salzgehaltszunahme, z. B. durch Verdunstung oder Eisbildung.

Nimmt die Dichte des Wassers mit der Tiefe rasch zu, bleibt die Wirkung der *thermohalinen Konvektion* auf eine relativ flache Oberschicht begrenzt (Abb. 10.3.1–7), wobei diese in Form einer Dichte-Sprungschicht gegen die Tiefe gesperrt wird. Das ist in tropischen und subtropischen Meeren ganzjährlich der Fall. Bei geringer Dichteschichtung, z. B. in den Polarmeeren, reicht die thermohaline Konvektion zum Teil bis zum Meeresboden. Neben der thermohalinen Konvektion wirkt die *dynamische Konvektion*, die durch äußere Kräfte erzwungene turbulente Vermischung benachbarter Wasserkörper.

10.3.1.3. Ozeanische Bewegungen. Meereswellen. Unter dem Einfluß äußerer Kräfte – Wind, Luftdruckunterschiede, Gravitationseinfluß von Sonne und Mond, stoßartige Anregung durch Erdbeben oder Gletscherkalben – entstehen an der Meeresoberfläche *Wellen* unterschiedlicher Längen, Höhen und Perioden (Tab. 10.3.1–8). Die (wie meist) durch Wind angeregten Wellen werden durch die Windgeschwindigkeit, die Streichlänge und die -dauer des Windes bestimmt. Beobachtungen aus dem Nordatlantik ergeben den in Abb. 10.3.1–9 enthaltenen mittleren Zusammenhang.

Über hinreichend tiefem Wasser (Wassertiefe $W \gg$ Wellenlänge L) führen die Wasserteilchen der Oberfläche ortsgebundene Kreisbahnen (*Orbitalbahnen*) aus. Bei den Windwellen ist die Wellenhöhe kleiner als der Durchmesser eines Rollkreises und ergibt eine verkürzte Zykloide oder Trochoide. Aus diesen Eigenschaften ergibt sich für die Wellenperiode $T = \sqrt{2\pi L/g} \approx 0{,}8\sqrt{L}$ und für die Ausbreitungsgeschwindigkeit (Phasengeschwindigkeit) $c = \sqrt{gL/(2\pi)} \approx 1{,}25\sqrt{L}$. Die Perioden solcher Trochoiden-

Tab. 10.3.1–8 Hauptarten der Meereswellen

Bezeichnung	Periode	Maximalhöhe
Windsee	bis 12 s	≧ 25 m
Dünung	10…30 s	einige m
Brandungsschwebungen	1…3 min	einige m
Tsunamis	10…30 min	bis 35 m
Seiches	min…Tag	1…2 m
Halbtagtiden	≈ 1/2 Mondtag	bis 13 m
Eintagstiden	≈ 1 Sonnentag	> 4 m

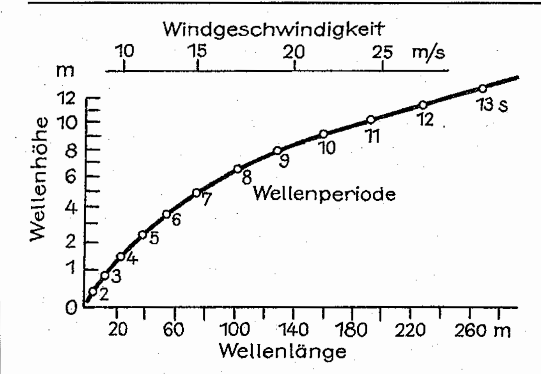


Abb. 10.3.1-9 Mittlerer Zusammenhang von Wellenlänge, -höhe und -periode mit der Windgeschwindigkeit im Nordatlantik (nach SCHUMACHER)

wellen hängen von der Wellenlänge ab, dabei wandern die längeren Wellen schneller als die kürzeren. Mit der Tiefe nimmt die Kreisbewegung der Teilchen ab; in der Tiefe einer Wellenlänge wird sie unbedeutend klein.

Wird die Wassertiefe W klein gegen die Wellenhöhe H, etwa bei $W < 2H$, setzt das *Branden* ein, eine Aufsteilung der Welle, schließlich ein Überbrechen des Wellenkammes. Dies tritt ein, wenn die Orbitalgeschwindigkeit im Wellenkamm die gebremste Ausbreitungsgeschwindigkeit der Welle übersteigt. Das ist dann der Fall, wenn das Verhältnis der Wellenhöhe zur Wassertiefe im Wellental Werte von 3/4 bis 4/5 erreicht.

Das natürliche Wellenbild zeigt als Interferenzerscheinung von Wellen wenig unterschiedlicher Periode oft quasiperiodisch besonders hohe Wellen. Das reale Wellenfeld ist das Ergebnis einer Überlagerung von Wellen unterschiedlicher Perioden.

An Trennschichten zweier Wasserkörper treten oft *interne Wellen* auf. Meist haben sie ihren Ursprung in Gezeitenerscheinungen. Sie können aus dem Entstehungsgebiet herauswandern und existieren mit sehr großer Variationsbreite ihrer Perioden und Amplituden. Sie können auch zu internen Brandungswellen führen.

Gezeiten. Die mit *Pegeln* gemessenen *Wasserstandsschwankungen*, Abweichungen von der idealen Ruhelage der Meeresoberfläche, können sehr vielfältige Ursachen haben: Dichteunterschiede, Stau von Meeresströmungen, Windstau, Flußeinmündungen u. a.

Die *Gezeiten* sind periodische Meeresspiegelschwankungen, die auf Grund der Gravitationswirkung von Sonne und Mond entstehen und durch die Topografie der Erde modifiziert werden. Besonders hoher *Tidenhub* (Wasserstandsdifferenz zwischen Niedrig- und Hochwasser) tritt dort auf, wo die Gezeitenströme sich in Buchten oder trichterförmigen Flußmündungen stauen. Die höchsten Gezeiten treten in der Fundybucht, Kanada, mit einem Tidenhub von 16 m zur Springzeit auf.

Die Gezeiten entstehen im Zusammenwirken von zwei Kräftegruppen: Gezeitenkräfte und Zentrifugalkräfte. Die Gravitationskraftwirkung auf ein Teilchen des festen Erdkörpers, der Hydro- oder der Atmosphäre ändert sich mit dem Abstand vom Mittelpunkt des Mondes und der Sonne. Außerdem resultieren aus dem Zweikörpersystem Erde–Mond wegen deren Rotation um einen gemeinsamen Schwerpunkt Fliehkräfte, die auf der Erde vom Mond weggerichtet und parallel zur Verbindungslinie beider Körper wirken. Dabei sind die lunaren Effekte wegen des relativ geringen Abstandes Erde–Mond mehr als doppelt so groß wie die solaren. So entstehen Überlagerungen von mond- und sonnentägigen Doppelschwingungen, deren resultierende Wirkung *Ebbe* und *Flut* mit *Spring-* und *Nipptiden* (maximale und minimale Amplitude) mit vielen regionalen Einflüssen und Besonderheiten hervorruft.

Meeresströmungen. Die im wesentlichen mit horizontalen Komponenten verlaufenden großräumigen Wasserbewegungen, die *Meeresströmungen*, erhalten ihren Antrieb von außen durch den Wind (*Triftströmungen*) oder durch Druckfelder, die im Meer in Verbindung mit den thermohalinen Strukturen aufgebaut werden (*Gradientströmungen*).

Triftströmungen. Der Wind übt auf die Wasseroberfläche über die *Schubspannung* eine Kraft aus. Durch turbulente Impulsübertragung von der oberen windbewegten Wasserschicht nach unten und infolge der ablenkenden Kraft der Erdrotation, der *Corioliskraft*, erhalten Triftströme ein charakteristisches Strömungsprofil, das vereinfacht, aber in guter Näherung an die Realität durch die Ekman-Spirale (Abb. 10.3.1–10) beschrieben wird. Dieses Modell geht von tiefenunabhängigen Koeffizienten des turbulenten Austausches aus und zeigt folgende Eigenschaften eines Triftstromes: Die Strömungsrichtung an der Oberfläche ist um 45° nach rechts (N-Halbkugel) von der Windrichtung abgelenkt und die Strömungsgeschwindigkeit beträgt ≈ 1,5 % der Windgeschwindigkeit. Mit zunehmender Tiefe beschreiben

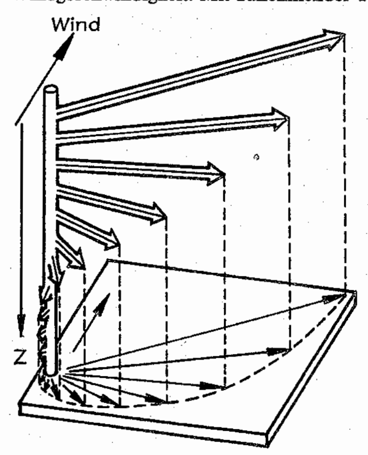


Abb. 10.3.1–10 Triftstromprofil (Ekman-Spirale)

die Triftstromvektoren eine Spirale. Die Tiefe, bei der der Triftstromvektor entgegengesetzt zu dem der Oberfläche zeigt und dabei einen sehr kleinen Betrag angenommen hat, gilt als *Reibungstiefe* oder *Tiefe der Reibungsschicht* (auch *Ekman-Schicht* genannt). Sie wächst mit dem turbulenten *Austauschkoeffizienten* und nimmt mit der Breite ab. Charakteristische Werte liegen bei ≈ 80 m (10° Br.) bis 40 m (70° Br.). Als Integral über das Triftstromprofil läßt sich der zugehörige Transport an Wassermasse bestimmen. Im offenen Meer ist stets eine Überlagerung von reinem Triftstrom und Gradientstrom vorhanden.

Gradientströmungen. Bei unausgeglichener Druckverteilung über einer Niveaufläche (der Schwere) wirkt in Richtung des Druckgefälles eine Kraft, die *Gradientkraft*, die zusammen mit der Corioliskraft die *Gradientströmung* im Meer

bestimmt. Im stationären Zustand und bei Vernachlässigung der Reibung und der Isobarenkrümmung erhält man eine Stromrichtung parallel zu den Isobaren mit einer Geschwindigkeit $c = (1/\varrho f)\,\partial p/\partial n$, mit $f = 2\omega \sin \varrho$ = Coriolisparameter, $\partial p/\partial n$ = Druckgradient senkrecht zu den Isobaren (vgl. 10.4.2.4.).

In einer Tiefe z setzt sich der horizontale Druckgradient aus den Inhomogenitäten des Dichtefeldes, einem tiefenabhängigen Term, und der von der Tiefe unabhängigen Neigung der Meeresoberfläche ϱ zusammen, $(\partial p/\partial n)_z = g\int_0^z \partial\varrho/\partial n\, dz - g\varrho(\varphi)(\partial\varphi/\partial n)$. Aus der Tiefenabhängigkeit des horizontalen Druckgradienten leitet sich das Profil der Gradientströmung $v = v(z)$ ab. Wird der Gradient zu Null, herrscht bewegungsloses Gleichgewicht; so existiert eine „*Nullfläche*", d. h., Isobarfläche und Niveaufläche fallen zusammen.

Die in den Ozeanen auftretenden Dichtesprungschichten (meist warme Oberschicht über kälterer Unterschicht) fallen bei stabiler Schichtung und Bewegungslosigkeit in der Niveau- und der Nullfläche zusammen. Zwischen bewegten Wassermassen unterschiedlicher Dichte erhalten sich quasistationäre geneigte Grenzflächen, die der Bedingung $\gamma = -(f/g)(\varrho_2 c_2 - \varrho_1 c_1)/(\varrho_2 - \varrho_1)$ genügen, wobei γ der Neigungswinkel ist und Index 1 die obere, Index 2 die untere Schicht kennzeichnen.

In der Dynamik der Meeresströmungen ist die Wirksamkeit der Kontinuitätsbedingung zu berücksichtigen. Das besagt, daß konvergierende oder divergierende horizontale Wasserbewegungen mit vertikalen Strömungskomponenten gekoppelt sind. Eine konvergente Strömung der Oberschicht hat im Konvergenzgebiet eine Abwärtsbewegung, und umgekehrt.

10.3.1.4. Allgemeine ozeanische Zirkulation. Unter *allgemeiner ozeanischer Zirkulation* ist die Gesamtheit des Systems der Meeresströmungen im mittleren Verlauf zu verstehen, mit horizontalen und auch vertikalen Komponenten. Die ozeanischen *Strömungsgeschwindigkeiten* liegen im Größenordnungsbereich 10^{-3} bis 10^2 cm/s. Diese allgemeine Zirkulation wird bestimmt durch die raum-zeitliche Verteilung der verursachenden Kräfte, durch terrestrische Bedingungen (Land-Meer-Verteilung, Erdrotation) und durch die Verteilung der dissipativen (Reibungs-)Kräfte. Die oberflächennahe ozeanische Zirkulation wird entscheidend durch die *atmosphärische Zirkulation* geprägt. So bilden sich ozeanische antizyklonale Wirbel im Gebiet der subtropischen Hochdruckzellen aus, deren Schwerpunkte allerdings auf der Nordhalbkugel nach Westen hin verschoben sind. (Abb. 10.3.1–11). Aus der Vielfalt der Strömungen sind die „Westrandströmungen" besonders eindrucksvoll, hier vor allem der *Golf-* und der *Kuroshiostrom*. Der Golfstrom wird mit Wasser aus dem Nordäquatorialstrom gespeist, das von der Straße von Florida bis Kap Hatteras als *Florida-*, anschließend bis südlich der Neufundlandbank als *Golfstrom* in markanter Bündelung und mäandrierend fließt und schließlich als sich auffächernder *Nordatlantikstrom* gegen Westeuropa strömt.

Die „Ostrandströmungen" haben als wichtige Begleiterscheinung Auftriebsgebiete von Tiefenwasser aus meist 200

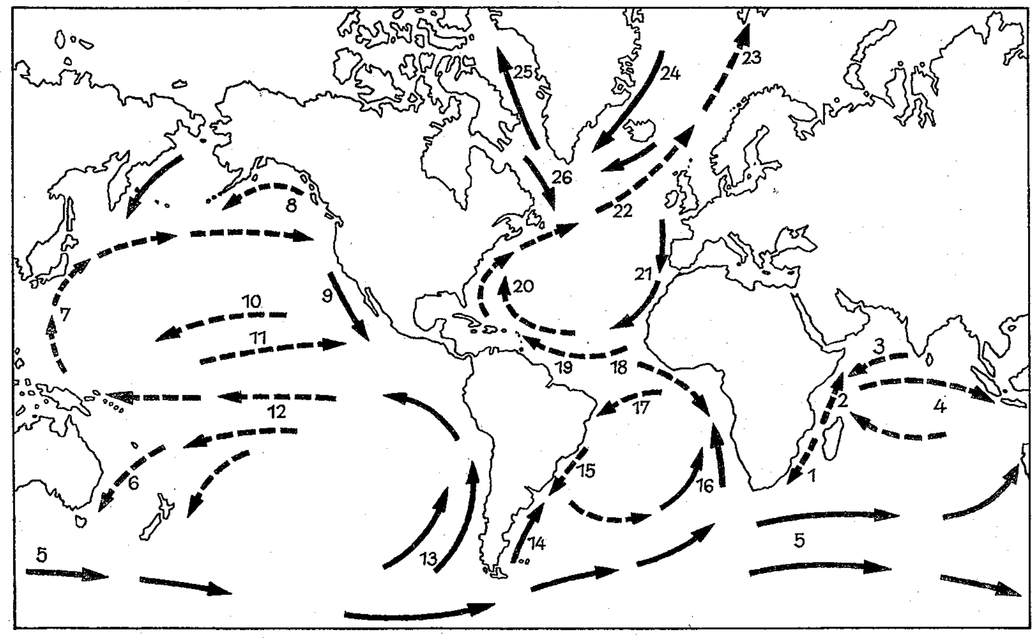


Abb. 10.3.1–11 Schema der Oberflächenströmungen der allgemeinen ozeanischen Zirkulation; 1 Agulhasstrom, 2 Somalistrom, 3 Nordäquatorialstrom, 4 Südäquatorialstrom, 5 antarktische Westwindtrift, 6 Ostaustralienstrom, 7 Kurushiostrom, 8 Nordpazifischer Strom, 9 Kalifornienstrom, 10 Nordäquatorialstrom, 11 Äquatorialgegenstrom, 12 Südäquatorialstrom, 13 Peru(Humboldt)strom, 14 Falklandstrom, 15 Brasilstrom, 16 Benguelastrom, 17 Südäquatorialstrom, 18 Äquatorialgegenstrom, 19 Nordäquatorialstrom, 20 Golfstrom, 21 Kanarenstrom, 22 Nordatlantikstrom, 23 Norwegenstrom, 24 Ostgrönlandstrom, 25 Westgrönlandstrom, 26 Labradorstrom; ⟶ Kaltwasserströme, – – –⟶ Warmwasserströme

bis 500 m Tiefe mit relativ niedriger Temperatur, geringerem Salz- und Sauerstoffgehalt und hoher Produktivität des Fischfanges.
Die ozeanische Tiefseezirkulation nimmt ihren Ausgang in den Meeren der hohen Breiten, wo Kaltwasser bis in große Tiefen unmittelbar in Wechselwirkung mit dem atmosphärischen Wärmehaushalt steht. Das hier gebildete Kaltwasser sinkt am Kontinentalabhang bis in große Tiefen ab und nimmt eine relativ stabile Lage ein. Die Tiefseezirkulation wird vorwiegend durch die Land-Meer-Verteilung beeinflußt, ist somit auf Nord- und Südhalbkugel unterschiedlich und wird auch durch die untermeerischen Rücken reguliert.

10.3.2. Physik der Binnengewässer

Nach der Verteilung des Wassers auf der Erde (Tab. 10.3.2–1) macht das *festländische Wasser* zwar nur einen kleinen Teil des Gesamtwassers aus, aber es tritt als Süßwasser auf, und seine Existenz und sein Kreislauf sind Voraussetzungen für festländische Lebensformen. Das festländische Wasser tritt als *Bodenfeuchte*, als *Grundwasser* und in verschiedenen Formen von *Oberflächengewässern* (Bäche, Flüsse, Seen) auf. Das als Eis festgelegte Wasser (*Kryosphäre*) bildet die *Gletscher* der Polargebiete und der Hochgebirge sowie das Eis des *Frostbodens* in Kaltklimaten, die „*Gefrornis*".
10.3.2.1. Grundwasser. Kapillare Hohlräume des Untergrundes leiten Sickerwasser in die Tiefe, das durch Stau über wasserundurchlässigen Schichten das Grundwasser bildet. Für den Boden (als wasserungesättigte Zone) ist eine unterschiedlich große spezifische Oberfläche ($n \cdot 10^{-2}$ m²/g, mit $1 < n < 10$) kennzeichnend. Mit zunehmendem Anteil der feineren Bestandteile (besonders Ton) nimmt die Oberflächenenergie zur Bindung des Wassers zu. Neben elektrostatischer Bindung wirkt vor allem die *Adhäsion* (kapillare Bindung). Als *Wassergehalt des Bodens* wird entweder das Verhältnis Masse Wasser zu Masse Boden oder Masse Wasser zu Volumen Boden angegeben.
Jede *Fließbewegung* im Boden ist eine Folge der Gesamtheit aller Kräfte (*hydraulischer Potentialgradient*), die auf das Wasser einwirkt. Im wassergesättigten Bereich gilt $v = -K\,\mathrm{d}h/\mathrm{d}s$ (*Gesetz von Darcy*), mit K = Durchlässigkeitsbeiwert (für Kies $> 10^{-1}$ cm/s, für Ton $< 10^{-5}$ cm/s), $\mathrm{d}h/\mathrm{d}s$ = hydraulischer Gradient.
Im ungesättigten Boden ist für K eine auch noch vom volumetrischen Wassergehalt W_v abhängige Größe wirksam.
Zum *Grundwasser* gehört alles Wasser, das die Hohlräume der Erdkruste zusammenhängend ausfüllt und nur dem hydrostatischen Druck unterliegt. Der Abfluß Q (Menge je Zeit) durch einen Grundwasserleiter ist darstellbar in der Form $Q = (K_g F/s)\,\Delta h_r$ (Δh_r = Reibungshöhe ≙ Reibungsbeiwert, K_g = Koeffizient in Abhängigkeit von Eigenschaften des Wasserleiters und des Wassers, F = Durchflußquerschnitt, s = Fließlänge).

Tab. 10.3.2–1 Süßwassermengen der Erde

	Wassermenge in km³	Anteil an der Gesamtwassermenge der Erde in %	Anteil an der Süßwassermenge in %
Gesamtsüßwassermenge der Hydrosphäre	28253200	1,94	100
Kryosphäre (Inlandeis und Gletscher)	24000000	1,65	85
Grundwasser (nur aktiver Austauschbereich)	4000000	0,275	14,15
Seen und Stauseen	155000	0,011	0,55
Bodenfeuchte	83000	0,006	0,29
in der Atmosphäre	14000	0,001	0,05
in Flüssen	1200	0,0001	0,004

10.3.2.2. Flüsse und Binnenseen. Zu den festländischen Oberflächengewässern gehören die *Fließgerinne* und die *Seen*. Erstere werden nach ihrer Wasserführung eingeteilt gemäß Tab. 10.3.2–2. Mit dem bewegten Wasser erfolgen Materialtransporte als *Geröllfracht*, *Schwebstofftransporte* oder *Gelöstes*. Die Flüsse sammeln den Oberflächenabfluß eines topografisch festgelegten Niederschlagsgebietes, wobei auch Grundwasseraustritte in die Flüsse einmünden. Die hydrologisch interessierenden Abflußwassermengen werden bestimmt durch Messungen der Fließgeschwindigkeitsverteilung über den Querschnitt des Flusses oder durch „*Überfallmessungen*" an Meßwehren.

Tab. 10.3.2–2 Einteilung der Oberflächengerinne nach der Wasserführung

Gerinne	Wasserführung in $m^3 s^{-1}$
Bäche	10...20
kleine Flüsse	20...200
große Flüsse	200...2000
Ströme	> 2000

Die Bedeutung der Seen liegt sowohl in ihrem hydrologischen Auffang- und Speichervermögen als auch in ihrer klimatischen Wirkung auf die Umgebung und in der Eigenart eines Biotops. Die maßgebliche Energiezufuhr erfolgt über die Sonnenstrahlung. Die Albedo von Wasserflächen ist stark vom Einfallswinkel der Strahlung abhängig und beträgt bei Sonnenhöhen 90° 30° 20° 10° entsprechend 2 %, 6 %, 13 % und 35 %. Die Strahlungsextinktion im Wasser hängt zusätzlich von der Trübung durch Suspensionen ab und beträgt etwa über Schichtdicken von 1 cm: 30 %, 10 cm: 50 %, 1 m: 65 %, 10 m: 80 %. In wenig durchmischten Seen entsteht dadurch ein vertikales Temperaturprofil mit relativ warmer Oberschicht und kaltem Tiefwasser; dazwischen liegt eine „*Sprungschicht*" mit sehr starkem Temperaturgradienten. Wegen der maximalen Wasserdichte bei 4 °C durchmischt sich ein See bei Abkühlung bis auf diese Temperatur, und es entsteht bei weiterer Abkühlung eine stabil geschichtete Oberflächenschicht mit Gefrierbereitschaft. Die gebildete Eisdecke wächst sehr langsam wegen der schlechten Wärmeleitfähigkeit des Eises.
10.3.2.3. Kryosphäre. *Meereis* entsteht durch die Absenkung des Gefrierpunktes mit dem Salzgehalt erst bei Abkühlung der Meeresoberfläche unter 0 °C (vgl. 10.3.1.1.). Die *Festlandsgletscher* bilden sich aus der Akkumulation fester Niederschläge über einen längeren Zeitraum unter Wirkung thermischer und druckbedingter Metamorphose des Schnees. Aus den unterschiedlichen Kristallformen des

Schnees entsteht über wechselndes Tauen und Gefrieren sowie durch dampfförmige Wasserverlagerungen, durch örtliche Temperatur- und Dampfdruckunterschiede Altschnee und Firn, durch weitere Verdichtung und Verringerung des Porenvolumens Firneis (mit $\varrho = 0{,}8 \ldots 0{,}85\,\mathrm{g/cm^3}$), schließlich transparentes Gletschereis ($\varrho = 0{,}91\,\mathrm{g/cm^3}$).

Bei den „kalten Gletschern" (der Polargebiete) liegt die Temperatur in allen Tiefen unter dem jeweiligen Druckschmelzpunkt. Die „temperierten Gletscher" (z. B. in den Alpen) haben ein vertikales Temperaturgefälle, das etwa der druckabhängigen Schmelzpunkterniedrigung entspricht. Sie führen ganzjährlich Schmelzwasser, das einen Wasser- und Massenverlust darstellt. Die Gesamtheit aller Masseverluste des Gletschers (Abfluß, Verdunstung, Winddrift des Firns) ist die *Ablation*. Zusammen mit der Akkumulation bestimmt sie den *Massehaushalt* des Gletschers, dessen Bewegung aufgrund der plastischen Eigenschaften des Eises durch diesen Massehaushalt und durch das Bodenrelief bestimmt wird.

10.3.3. Haushalt und Kreislauf des Wassers

Der *Wasserhaushalt* ist die quantitative Bilanz der Komponenten des *Wasserkreislaufes*. Er besteht aus den drei Komponenten *Niederschlag* N, *Verdunstung* V und *Abfluß* A, $N = V + A$, wobei sich der Abfluß in den oberirdischen und den unterirdischen aufgliedert. Für kurzzeitige Wasserhaushaltsbetrachtungen ist noch die Änderung der Bodenspeicherung (Rücklage oder Aufbrauch) und die im Bewuchs festgehaltene Niederschlagsmenge (Interzeption) zu berücksichtigen.

Während der Niederschlag und der Abfluß trotz vieler Probleme, die die Genauigkeit der Bestimmung von Gebietsniederschlägen oder -abflüssen beeinträchtigen, hinreichend sicher meßbar sind, ist die direkte Messung der Verdunstung (*Evaporation*) kompliziert; sie wird deshalb meist indirekt aus der Summengleichung des Wasserhaushaltes ermittelt. Über Seen ist die Evaporation V_s über eine empirische Gleichung $V_s = a\,v^b\,(e_1 - e_2)$ abschätzbar, wobei v die Windgeschwindigkeit in 2 m Höhe, $e_1 - e_2$ der Dampfdruckunterschied zwischen Seeoberfläche und 2 m Höhe und a, b Koeffizienten sind, die individuelle Eigenschaften eines Sees beinhalten und speziell bestimmt werden müssen. Mit *Lysimetern*, unterschiedlich großen Behältern für den Boden, mit oder ohne Bewuchs, sind die einzelnen Komponenten des Wasserhaushaltes unter den jeweiligen Bedingungen direkt bestimmbar. Die mittlere Verdunstung liegt in Mitteleuropa (mit Niederschlägen um 650 mm/Jahr) bei 500 mm/Jahr.

Im *Wasserkreislauf* für die gesamte Erde (Abb. 10.3.3–1) steht einem atmosphärischen Wassergehalt von $\approx 12\,300\,\mathrm{km^3}$ Wasser ein jährlicher Gesamtniederschlag von $\approx 475\,000\,\mathrm{km^3}$ Wasser gegenüber; der atmosphärische Wasserdampf wird also jährlich fast 40mal erneuert. Antriebskraft für diesen Kreislauf ist die Sonnenernergie. Über die Verdunstung wird als latenter Wärmestrom (vgl. 10.4.1.3.) der Atmosphäre eine Energiemenge von $\approx 10{,}7 \cdot 10^{23}$ J/Jahr zugeführt, das entspricht einer mittleren globalen Leistungsdichte des latenten Wärmestromes von 67 W/m².

10.4. Physik der Atmosphäre

10.4.1. Wärmehaushalt der Atmosphäre

10.4.1.1. Solare Einstrahlung und optische Eigenschaften der Atmosphäre. Die mit der Strahldichte (= Solarkonstante, vgl. 10.1.5.) von $\bar{I}_0 = 1\,396\,\mathrm{W/m^2}$ an der Obergrenze der Atmosphäre verfügbare Sonnenstrahlung mit einer *spektralen Energieverteilung* gemäß Tab. 10.4.1–1 unterliegt beim Durchgang durch die Atmosphäre einer Schwächung (*Extinktion*) nach dem *Bouguer-Lambertschen Gesetz* ($(I_\lambda/I_{0\lambda}) = \mathrm{e}^{-a_\lambda m}$ (a_λ in cm^{-1} = Extinktionskoeffizient, m = durchstrahlte Luftmasse = reduzierte Weglänge in der Atmosphäre mit konstanter Normaldichte). Diese Extinktion schließt die Streuung (an Luftmolekülen, Tröpfchen und Teilchen) und die selektive Absorption durch Atmosphärenbestandteile ein. Die Streuung an den (sehr kleinen) Luftmolekülen erfolgt nach dem *Rayleigh-Gesetz*, nach dem

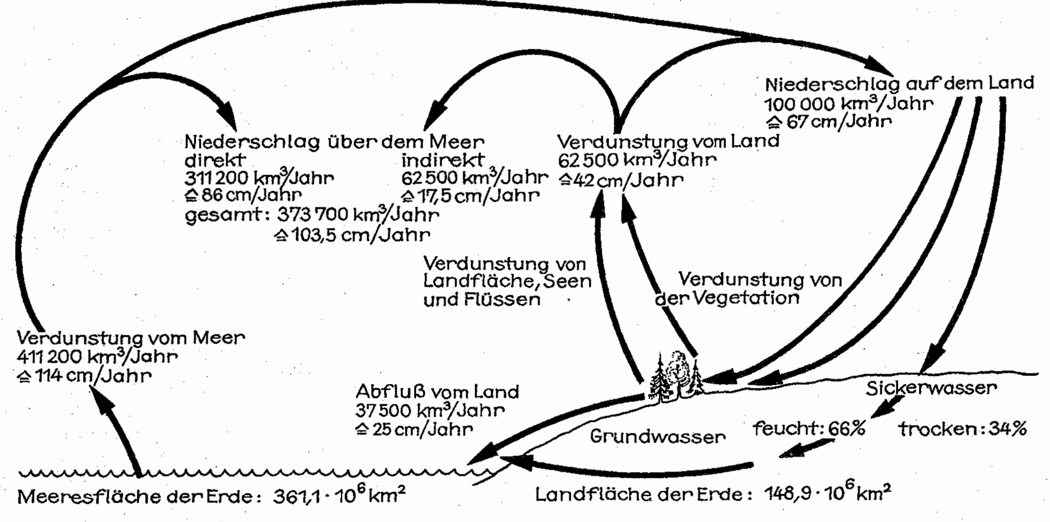


Abb. 10.3.3–1 Schema des globalen Wasserkreislaufes (nach MARCINEK)

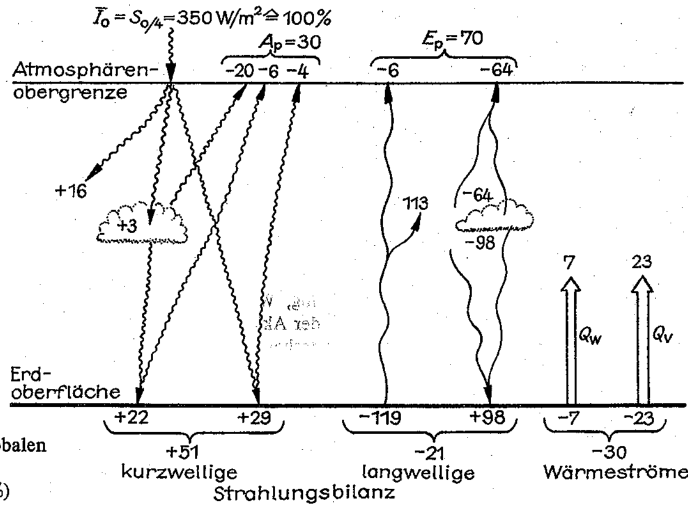


Abb. 10.4.1-3 Schema des mittleren globalen Wärmehaushaltes der Nordhalbkugel (Zahlenangaben: Relativwerte zu $\bar{I}_0$ in %)

ein Extinktionskoeffizient $a_{R,\lambda} = 8{,}79 \cdot 10^{-3} \lambda^{-4{,}09}\ \text{cm}^{-1}$ gilt, d.h., die Streuung ist besonders groß bei kurzen Wellenlängen. Das erklärt das Himmelsblau einer sauberen Atmosphäre und auch die Rotfärbung bei horizontnaher Sonne, weil dann der extingierte Anteil der Farbe Blau besonders groß ist und Rot in der direkten Sonnenstrahlung überwiegt. Die Streuung einer Dunstatmosphäre mit einem Kollektiv verschieden großer Teilchen erfolgt nach dem *Mie-Streuungsgesetz*. Im Mittel gilt ein Extinktionskoeffizient $a_{D\lambda} = \beta_\lambda^{-1,3}$, wobei β ein Faktor der Trübungsintensität ist. Für die selektive Absorption sind im Bereich des Sonnenspektrums besonders die atmosphärischen Bestandteile Wasserdampf, Kohlendioxid und Ozon (O_3) verantwortlich. Durch das Ozon gelangen Wellenlängen unter 300 nm nicht bis zur Erdoberfläche.

An der Erdoberfläche trifft die modifizierte Sonnenstrahlung als direkte Sonnen- I und diffuse Himmelsstrahlung D ein, deren Summe die *Globalstrahlung* G ergibt, die Strahldichte auf die Flächeneinheit der Erdoberfläche, $G = I \cos z + D$ (z = Zenitdistanz der Sonne). Bei bewölktem Himmel enthält die Streustrahlung auch die diffus durch die Wolken dringende Strahlung und die Reflexstrahlung beleuchteter Wolkenflächen.

Durch Refraktion, Reflexion und Streuung kann in der Atmosphäre eine Vielfalt optischer Erscheinungen bei speziellen Schichtungsstrukturen oder durch Tröpfchen oder Eiskristalle auftreten, wie *Halobildungen*, *Regenbögen*, *Kränze* und *Glorien*.

Tab. 10.4.1-1 Spektrale Energieverteilung der Sonnenstrahlung außerhalb der Erdatmosphäre bei mittlerer Sonnenentfernung

λ in µm	$I_{0,\lambda}$ in W/m² µm	λ in µm	$I_{0,\lambda}$ in W/m² µm
0,225...0,3	17	1,0...1,1	66
0,3...0,4	110	1,1...1,2	55
0,4...0,5	200	1,2...1,3	45
0,5...0,6	193	1,3...1,5	66
0,6...0,7	162	1,5...2,0	84
0,7...0,8	128	2,0...3,0	54
0,8...0,9	101	3,0...11,0	27
0,9...1,0	81	11,0...30,0	0,7

Tab. 10.4.1-2 Reflexionsvermögen natürlicher Oberflächen für solare (A_s) und terrestrische Strahlung (A_t)

	A_s	A_t
Schneedecke	0,5...> 0,9	~ 0,005
Sand	0,25...0,45	~ 0,1
Rasen	0,15...0,25	~ 0,016
Äcker	0,05...0,15	~ 0,1
Wasserflächen	0,03...0,1	~ 0,04
Wolken	0,5...0,8	0...0,1

10.4.1.2. Atmosphärischer Strahlungshaushalt. Unter dem *Strahlungshaushalt* wird die Gesamtheit der Strahlungskomponenten der *solaren Einstrahlung* und der *terrestrischen Ausstrahlung* zusammengefaßt. Aus der Globalstrahlung G wird je nach dem Reflexionsvermögen A_s natürlicher Oberflächen (Tab. 10.4.1-2) ein Anteil $Q_s = (1 - A_s)\, G$ absorbiert. Erdoberfläche und auch Atmosphäre senden Eigenstrahlung (*terrestrische Strahlung*) gemäß dem *Stefan-Boltzmann-* und dem *Kirchhoff-Gesetz* aus, wonach gilt: $B_t \sigma T^4 = (1 - A_t)\, \sigma T^4$ (B_t = Absorptions- bzw. Emissionsvermögen, A_t = Reflexionsvermögen für terrestrische Wärmestrahlung). Diese Größen sind wellenlängenabhängig und werden für integrale energetische Betrachtungen als gewichtete Mittelwerte verwendet (vgl. Tab. 10.4.1-2). Schnee hat z.B. hohe Reflexion im sichtbaren Strahlungsbereich und ist für Wärmestrahlung fast „schwarz".

Der terrestrische Strahlungsbereich liegt zwischen $\lambda = 4$ und 100 µm mit einem Maximum bei $\lambda = 10$ µm, was einer Strahlungstemperatur von ≈ 288 K entspricht. Für die von der Erdoberfläche ausgehende Wärmestrahlung („*Ausstrahlung*") ist die Atmosphäre durch ihre Bestandteile Wasserdampf und Kohlendioxid ein wirksamer Absorber. Die wesentlichen Absorptionsbanden liegen für H_2O in den Bereichen $\lambda = 5$ bis 8 µm und $\lambda > 17$ µm, für CO_2 besonders

im Gebiet $\lambda = 13$ bis 17 µm. Es existiert also ein Durchlaßbereich („Fenster“) für langwellige Strahlung zwischen 8 und 13 µm Wellenlänge. Nach dem Kirchhoff-Gesetz emittieren die atmosphärischen Schichten in gleicher Art wie sie absorbieren, so daß atmosphärische Wärmestrahlung teils in den Raum, teils als *Gegenstrahlung* zur Erdoberfläche zurückgeht („Glashauseffekt“). Als Differenz zwischen Ausstrahlung der Erdoberfläche und Gegenstrahlung entsteht die *effektive Ausstrahlung*, die somit sehr wesentlich durch das atmosphärische CO_2 und durch den (sehr variablen) Wasserdampfgehalt und durch die Bewölkung mitbestimmt wird.

10.4.1.3. Wärmehaushalt der Erdoberfläche und der Atmosphäre. Der gesamte *Wärmehaushalt* an der Erdoberfläche setzt sich aus dem Strahlungshaushalt Q_S und der Bilanz der Wärmeströme zusammen. Zu letzterer zählt, abgesehen von Komponenten sehr viel kleinerer Wirksamkeit, wie dem Erdwärmestrom, der turbulente *Strom fühlbarer Wärme* Q_W und der turbulente *Strom latenter Wärme* Q_V, der über den Wasserdampf die aufgenommene Verdunstungswärme latent transportiert, bis sie durch Kondensation freigesetzt und der Atmosphäre zugeführt wird. Dieser Wärmehaushalt ist orts- und zeitabhängig und kommt im *Klima* des jeweiligen Ortes der Erdoberfläche zum Ausdruck. Im globalen Mittel und im Jahresdurchschnitt haben die einzelnen Komponenten die aus dem Schema der Abb. 10.4.1–3 ersichtlichen Anteile.

Zur Summengleichung des Wärmehaushaltes an der Erdoberfläche $Q = Q_S + Q_W + Q_V$ kommt auf der Meeresoberfläche noch das Wärmetransportglied der Wasserbewegungen hinzu. In der Atmosphäre greift ebenfalls der horizontale Luftmassentransport (*Advektion*) stark in den Wärmehaushalt eines ortsfesten Punktes ein. Diese ozeanischen und atmosphärischen Wärmetransporte sorgen für eine Milderung der meridionalen Temperaturunterschiede, wie sie sich bei reinem *„Strahlungsklima“* ausbilden würden.

Tab. 10.4.2–1 Wichtige Größen der Physik der Atmosphäre

virtuelle Temperatur $T_v = T(1 + 0{,}608\, s)$ in K

(T_v ist die Temperatur, die trockene Luft hat, um unter demselben Druck dieselbe Dichte zu haben wie feuchte Luft mit der spezifischen Feuchte s; daraus ergibt sich die Gasgleichung für feuchte Luft: $p/\varrho = R_L T_v$).

potentielle Temperatur $\theta = T\left(\frac{1\,000\,[\text{hPa}]}{p}\right)^k$ mit $k = \frac{c_p - c_v}{c_p} = 0{,}285$

(θ ist die Temperatur, die Luft des Ausgangszustandes p, T bei adiabatischer Verschiebung in das Niveau 1 000 hPa annimmt).

Partialdruck des Wasserdampfes (Dampfdruck) e in hPa

maximal möglicher Dampfdruck

über Wasser $E_w = 6{,}107 \cdot 10^{\frac{7,5\,t}{237+t}}$ in hPa

über Eis $E_e = 6{,}107 \cdot 10^{\frac{9,5\,t}{265,5+t}}$ in hPa

relative Luftfeuchte $RF = (e/E) \cdot 100$ in %

Mischungsverhältnis $m = \varrho_W/\varrho_L = e(R_L/\varrho_L R_W)$ $= 0{,}622\, e/(p - e)$ in g H_2O/g Luft (Index W = Wasserdampf, L = trockene Luft)

spezifische Feuchtigkeit $s = \varrho_W/(\varrho_L + \varrho_W)$ $= 0{,}622\, e/(p - 0{,}378\, e) \approx 0{,}622\, e/p$ in g H_2O/g feuchte Luft

absolute Feuchtigkeit (Dampfdichte) $a = \varrho_W = e/(e/R_W T)$ in g/cm³

Taupunkt $T_d = T(e = E)$ in K oder °C

Gaskonstante

für trockene Luft $R_L = 2{,}870\,5 \cdot 10^6$ in cm²/s²K

für Wasserdampf $R_W = 4{,}615\,2 \cdot 10^6$ in cm²/s²K

10.4.2. Grundzüge wetterbildender Prozesse

10.4.2.1. Barometrische Höhenbeziehungen. Die Verknüpfung von Luftdruck, Temperatur und Höhe wird durch die *statische Grundgleichung* beschrieben, $dp = -g\varrho dz = -g(p/R_L T_v)\,dz$ (Zusammenstellung der physikalischen Größen in Tab. 10.4.2–1). Da $g = g(\varphi, z)$ und $\varrho = \varrho(z)$ ist, erweist es sich als zweckmäßig, $g(\varphi, z) = -d\Phi$ und $\Phi - \Phi_0 = -\int_0^z g(\varphi, z)\,dz$ zu setzen. Φ ist dann die Arbeit, die geleistet werden muß, um eine Masseneinheit gegen die Schwerkraft über die Höhe z zu heben, das *Geopotential* in m^2s^{-2}. Legt man wegen der besseren Anschaulichkeit den 9,8ten Teil von Φ als 1 *geopotentielles Meter*, 1 gpm, fest und bezeichnet die zugehörige Höhenkoordinate mit h, so gilt: $g(\varphi, z)\,dz = g_n dh$ (g_n = Normalschwere). Die Integration der statischen Grundgleichung liefert für den fiktiven Fall einer homogenen Atmosphäre mit $\varrho = \varrho_0 = \text{const.}$ eine Gesamthöhe dieser Atmosphäre $H = p_0/g_n\varrho_0$, wofür man, von den Normalbedingungen an der Erdoberfläche ausgehend, $H \approx 8$ km erhält. Ersetzt man ϱ durch die Gas-

Tab. 10.4.2–2 Die Normal-(Standard-)Atmosphäre (nach der ICAO-Atmosphäre)

p in hPa	h in gpm	T in °C	p in hPa	h in gpm	T in °C
1 013,25	0	15,00	70	18 442	−56,50
1 000	111	14,28	60	19 419	−56,50
950	540	11,49	50	20 575	−56,50
900	988	8,58	40	21 991	−56,50
850	1 457	5,53	30	23 815	−56,50
800	1 949	2,33	20	26 398	−52,31
750	2 466	− 1,03	15	28 283	−46,65
700	3 012	− 4,58	10	31 020	−38,44
600	4 206	−12,34	9	31 747	−36,26
500	5 574	−21,23	8	32 568	−33,80
400	7 185	−31,70	7	33 509	−30,97
300	9 164	−44,57	6	34 609	−27,11
200	11 784	−56,50	5	35 930	−23,71
150	13 608	−56,50	4	37 575	−18,78
100	16 180	−56,50	3	39 745	−12,27
90	16 848	−56,50	2	42 896	− 2,81
80	17 595	−56,50	1	48 447	+ 9,50

gleichung, dann entsteht die Beziehung $\ln(p_2/p_1) = -(g_n/R_L)\int_{h_1}^{h_2}[1/T_v(h)]\,dh = -(g_n/R_L\bar{T}_v)(h_2 - h_1)$ (h_1 = untere, h_2 = obere Begrenzung der Luftsäule, p_1, p_2 = zugehörige Drücke, $\bar{T}_v$ = virtuelle Schichtmitteltemperatur). Mit Hilfe dieser Beziehung können Luftdruckreduktionen auf ein Referenzniveau, Höhenbestimmungen aus Druck- und Temperaturmessungen und Schichtmitteltemperaturen zwischen Flächen konstanten Druckes aus deren geometrischem Abstand bestimmt werden.

Die international vereinbarte *Normalatmosphäre* (Tab. 10.4.2–2) repräsentiert einen mittleren Zustand der Verknüpfungen p, T, h.

10.4.2.2. Adiabatische Prozesse und Schichtungsstabilität. Relativ rasch verlaufende atmosphärische Vorgänge können thermodynamisch als *adiabatische Vorgänge* (ohne Zufuhr oder Abgabe von Wärme) behandelt werden. Aus dem 1. Hauptsatz $dq = c_p\,dT - v\,dp = 0$ (für adiabatische Prozesse) erhält man den interessierenden Wert für eine Temperaturabnahme bei adiabatischer Vertikalbewegung unter Verwendung der Gasgleichung und der statischen Grundgleichung:

$$dT/dh = (v/c_p)dp/dh = (R_LT/c_pp)dp/dh$$
$$= (1/c_p\varrho)dp/dh = -g_n/c_p = -\gamma^*$$

Dieser *adiabatische Temperaturgradient* beträgt $\gamma^* = 9{,}76 \approx 10$ K/gpkm. Er gilt auch für wasserdampfhaltige Atmosphäre, da sich c_p für reine und für wasserdampfhaltige Luft nur um weniger als 1 % unterscheiden.

Tritt Wasserdampfsättigung ein, dann wird Verdampfungswärme frei und führt zu einem *feuchtadiabatischen Gradienten* γ_f^*, der eine Funktion des Druckes und der Temperatur ist.

Eine atmosphärische Schichtung wird dann als stabil bezeichnet, wenn ein aus der Ruhelage gebrachtes Masseteilchen sich wieder in die Ausgangslage zurückbewegt. Das ist dann der Fall, wenn es bei Verschiebung nach oben kälter, bei Verschiebung nach unten wärmer als seine Umgebung ist. Ist $\gamma = -\partial T/\partial h$ der vertikale Temperaturgradient in einer realen Luftsäule, so gilt unter Beachtung der Definition der potentiellen Temperatur (vgl. Tab. 10.4.2–1) für

Stabilität: $\gamma < \gamma^*$ und $\partial\vartheta/\partial h > 0$
Labilität: $\gamma > \gamma^*$ und $\partial\vartheta/\partial h < 0$
Indifferenz: $\gamma = \gamma^*$ und $\partial\vartheta/\partial h = 0$.

Analoges gilt für die *„Feuchtstabilität“* in bezug auf γ_f^* im Falle einer wasserdampfgesättigten Atmosphäre.

Neben mehreren anderen *Stabilitätsmaßen* wird häufig die „Richardson-Zahl“ $Ri = (g/T)(\partial\vartheta/\partial h)/(\partial v/\partial h)^2$ verwendet, die das Verhältnis von Stabilität zur kinetischen Energie der Windscherung je Masseneinheit angibt; $Ri > 0$ bei stabiler Schichtung usw.

In der freien Atmosphäre kann der labile Zustand nie auftreten, weil die Atmosphäre auf Labilisierung sofort mit einer Umlagerung in einen mindestens indifferenten Zustand reagiert. Innerhalb einer dünnen atmosphärischen Bodenschicht über stark aufgeheiztem Untergrund können zeitweilig „überadiabatische“ Temperaturgradienten auftreten. Dann vollzieht sich eine permanente Umlagerung in Form starker Turbulenz, die auch optisch als *„Luftflimmern“* sichtbar wird.

10.4.2.3. Wolken- und Niederschlagsbildung. Der maximal mögliche *Wasserdampfgehalt der Luft E* ist abhängig von der Temperatur (vgl. Tab. 10.4.2–1). Die E zugehörige Temperatur ist der *Taupunkt* (auch Taupunktstemperatur) T_d. Wird E über-, bzw. T_d unterschritten, kondensiert oder sublimiert soviel Wasserdampf, daß die Relation $e = E$ bzw. $T = T_d$ hergestellt wird. Kondensationsformen am Boden werden als *Tau* ($T > 0$ °C) oder *Reif* ($T < 0$ °C) bezeichnet. *Nebel* ist eine Wolkenart in Bodennähe. Beim Überschreiten des *Sättigungsdampfdruckes E* bildet sich eine entsprechend hohe Konzentration sehr kleiner Tröpfchen, die die Transparenz der Atmosphäre für Licht herabsetzen. Es gibt zahlreiche Entstehungsursachen: Wasserdampfzufuhr, und/oder Abkühlung durch Ausstrahlung, durch Mischung mit kälterer Luft, durch Wärmeabgabe an die Unterlage oder (adiabatisch) durch Druckerniedrigung.

Wolken bilden sich fast ausschließlich durch adiabatische Abkühlung infolge Vertikalbewegung, die bei wenig stabiler Schichtung in Form zellenförmiger Konvektion (mit relativ hohen Vertikalgeschwindigkeiten und der Bildung von Haufen- oder Cumuluswolken), bei stabiler Schichtung durch Gleitbewegungen längs Flächen konstanter potentieller Temperatur stattfinden (mit sehr langsamer Hebung und der Bildung von Schicht- oder Stratuswolken). *Wolkenbildung* setzt bei Hebung dann und dort ein, wo ein Luftpaket, das in Bodennähe die Temperatur T_0 und die spezifische Feuchte s_0 hat, unter adiabatischer Temperaturänderung ein Niveau p, T erreicht, in dem $s_0 = s_{max}$ wird (= *Kondensationsniveau*). Bei Erwärmung der bodennahen Luftschicht beginnt Wolkenbildung dann, wenn eine „Auslösetemperatur“ T_a am Boden erreicht wird. Mit dieser wird ein adiabatisches Temperaturgefälle zwischen Erdoberfläche und Kondensationsniveau hergestellt (Abb. 10.4.2–3). Über die vertikale Mächtigkeit der Wolke entscheidet die Temperaturstruktur oberhalb der Kondensationshöhe. Da bei weiterer Aufwärtsbewegung des nunmehr wasserdampfgesättigten Luftpaketes die Temperaturänderung feuchtadiabatisch verläuft, erstreckt sich die Wolke über den Höhenbereich oberhalb der Kondensationshöhe, in dem $\gamma^* > \gamma_f^*$ ist.

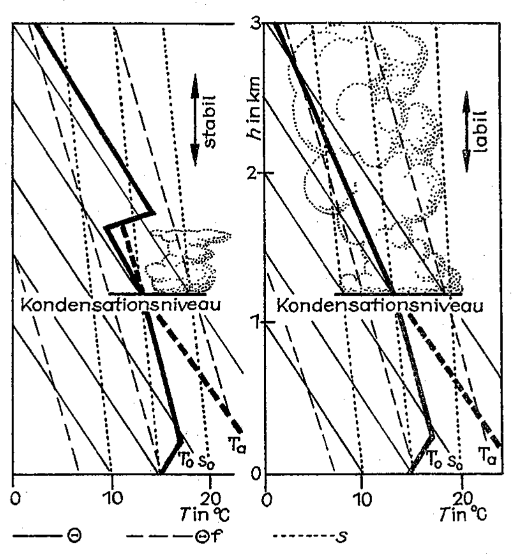


Abb. 10.4.2–3 Vertikale Zustandsverteilung bei Konvektionswolkenbildung; *fette Linie* vertikale T-Verteilung, *fette gestrichelte Linie* T-Verlauf in der Konvektionszelle, die sich nach Erreichen der Auslösetemperatur T_a bildet

Föhn, ein sehr warmer und trockener Wind auf der Leeseite von Gebirgen, entsteht aus der adiabatischen Abkühlung mit Wolken- und Niederschlagsbildung beim Überströmen des Gebirges. Fällt kein Niederschlag aus, verlaufen die adiabatischen Temperaturänderungen auf beiden Seiten des Gebirges spiegelbildlich. Fällt aber Niederschlag, setzt trockenadiabatische Erwärmung der Luft auf der Leeseite bereits oberhalb der Kondensationshöhe der Luvseite ein, so daß in gleichen Höhen (Druckniveaus) im Lee höhere Temperaturen herrschen.

Die Entstehung der *Wolkentröpfchen* und ihre Vergrößerung zu ausfallenden Tropfen oder Eiskörnern unterliegt neben den genannten makroskopischen Prozessen auch komplizierten mikrophysikalischen. Die Bildung eines Tröpfchens, die Zusammenlagerung von Wassermolekülen, erfordert eine Dampfspannung im Raum, die über dem normalen Sättigungsdampfdruck liegt. Das ist notwendig wegen der Oberflächenspannung der Flüssigkeit (vgl. Thomsonsche Formel). Unterstützt wird die Tropfenbildung in der Natur durch elektrische Ladungsträger und durch kleine Materieteilchen, die benetzbar oder wasserlöslich sind (*Kondensationskerne*). Dadurch sind für die erste Stufe der Tröpfchenbildung nicht so hohe Übersättigungen erforderlich, wie aus der Thomson-Formel hervorgeht, sondern nie mehr als 1 %.

Da der Oberflächendampfdruck über größeren Tröpfchen geringer ist als über kleinen, erfolgt ein Dampfstrom zu größeren Tröpfchen, die auf Kosten der kleinen anwachsen. Wolkentröpfchen haben einen Radius von $\approx 10^{-3}$ cm, große Regentropfen von $>10^{-1}$ cm; ein Wolkentröpfchen muß also auf das 10^6fache Volumen anwachsen, um ausfallen zu können. Es müssen zu der erfahrungsgemäß relativ raschen Bildung von Regentropfen andere Prozesse beitragen. Diese sind besonders mechanische Anlagerungsprozesse durch die Tropfenbewegung in der turbulenten Luftströmung der Wolke (*Koagulation*).

Graupel und *Hagel* sind am starken Aufwindfeld einer Schauer- oder Gewitterwolke mehrfach zwischen den Temperaturniveaus über und unter dem Gefrierpunkt wechselnde Kondensationsprodukte, bei denen die wiederholte äußere Wasseranlagerung gefriert und zu einer Schalenstruktur des Hagelkornes führt.

10.4.2.4. Luftbewegungen, Zirkulation. Sieht man von kleinräumigen, zellularen Luftbewegungen infolge zeitweiliger Schichtungslabilisierung ab, werden die (großräumigen) Luftbewegungen durch Beschleunigungen bestimmt, die sich aus dem Druckfeld ergeben, $\mathrm{d}\boldsymbol{v}/\mathrm{d}t = -(1/\varrho)\nabla p$. Legt man für die praktische Handhabung dieser Beziehung den Gradienten der Höhenlage (des Geopotentials) einer Isobarfläche im n, h-System zugrunde ($[\partial h/\partial n]_{p=\mathrm{const.}}$ mit h = Höhe in gpm), so ergibt sich mit $1/\varrho = -g_n(\partial h/\partial p)$: $\mathrm{d}\boldsymbol{v}_{\mathrm{hor.}}/\mathrm{d}t = g_n(\partial p/\partial n)/(\partial p/\partial h) = -g_n(\partial h/\partial n)_p = \nabla_{\mathrm{hor.}}\Phi(p)$.

Nach dieser Beziehung bewirken mit einem Temperaturfeld zusammenhängende Druckfelder ausgleichende *Luftbewegungen*, *Zirkulationen*. Eine *Land-See-Wind*zirkulation entsteht z. B. durch tagsüber stärkere Erwärmung über Land, was in der Höhe zu höherem Luftdruck, dadurch dort zu Luftmassenbewegung seewärts führt. Dieser Massenverlust erzeugt am Boden geringeren Luftdruck, der den auflandigen Seewind zur Folge hat. Nachts kehrt sich dieses System infolge des veränderten Temperaturfeldes um.

Auf großräumige Luftbewegungen wirkt die Erdrotation über ihre ablenkende Kraft, die *Corioliskraft*, ein. Ihre horizontale Komponente $C_{\mathrm{hor.}} = -2\Omega \sin\varphi\, \boldsymbol{k} \times \boldsymbol{v}_{\mathrm{hor.}} = -f\boldsymbol{k} \times \boldsymbol{v}_{\mathrm{hor.}}$ mit $f = 2\Omega\sin\varphi$ = Coriolisparameter steht senkrecht auf der Bewegungsrichtung, nach rechts auf der Nord-, nach links auf der Südhalbkugel. Durch diese Kraft wird keine Arbeit geleistet; sie ist eine Scheinkraft, die proportional zur Geschwindigkeit wirkt, aber die Geschwindigkeit selbst nicht ändert. Sie hängt von der geografischen Breite φ ab.

Die allgemeine *Bewegungsgleichung* auf der rotierenden Erde lautet $\mathrm{d}\boldsymbol{v}/\mathrm{d}t = -(1/\varrho)\nabla p - 2\boldsymbol{\Omega} \times \boldsymbol{v} + \mathrm{g} + (1/\varrho)F$, wobei $(1/\varrho)F$ die Gesamtheit aller weiteren Beschleunigungen, z. B. die als negative Beschleunigung wirkende Reibung enthalten soll. Für die Horizontalkomponente wird daraus $\mathrm{d}\boldsymbol{v}_{\mathrm{hor.}}/\mathrm{d}t = -(1/\varrho)\nabla_h p - f\boldsymbol{k} \times \boldsymbol{v}_{\mathrm{hor.}} + (1/\varrho)F$. Als einfachsten Fall des Gleichgewichtes zwischen Wind- und Druckfeld unter beschleunigungs- und reibungsfreien Bedingungen ergibt sich für geradlinig verlaufende Isobaren $0 = (1/\varrho)\nabla p + f\boldsymbol{k} \times \boldsymbol{v}$ und daraus $\boldsymbol{v}_{\mathrm{geostr.}} = \boldsymbol{k} \times (1/\varrho f)\nabla p$ oder $v_{\mathrm{geostr.}} = -(1/f)(\partial\Phi/\partial n)_p$.

Dieser *„geostrophische Wind“* weht infolge der Coriolisbeschleunigung mit einer dem Druckgradienten proportionalen Geschwindigkeit senkrecht zum Druckgradienten, also parallel zu den Isobaren und zwar so, daß auf der Nordhalbkugel der höhere Druck rechts, der tiefere links von der Strömungsrichtung liegt (Abb. 10.4.2–4a).

Führt man in diese geostrophische Windgleichung als Beschleunigungsterm nur die Zentrifugalbeschleunigung ein, die infolge gekrümmter Isobaren wirkt, erhält man in skalarer Schreibweise als Beziehung für diesen (reibungsfreien!) *Gradientwind* (mit r = Krümmungsradius der Isobaren): $\pm(v^2/r) + fv + (\partial\Phi/\partial n)_p = 0$, wobei das positive Vorzeichen im 1. Term für zyklonal gekrümmte Isobaren gilt (*zyklonal* = Krümmung um den Ort mit tieferem Luftdruck); das negative Vorzeichen gilt entsprechend für ein antizyklonales Druckfeld (Abb. 10.4.2–4b). Man erkennt, daß unter sonst gleichen Bedingungen des Druckfeldes die Windgeschwindigkeit bei antizyklonaler Krümmung höher ist. Daß es sich in der beobachteten Windstärkestatistik anders verhält, liegt an den im Durchschnitt größeren Druckgradienten in Tiefdruckgebieten.

Treten weitere Beschleunigungen auf, z. B. in Druckfeldern mit nichtparallelen Isobaren, ergeben sich infolge der Massenträgheit der Luft *ageostrophische Abweichungen*, d. h.

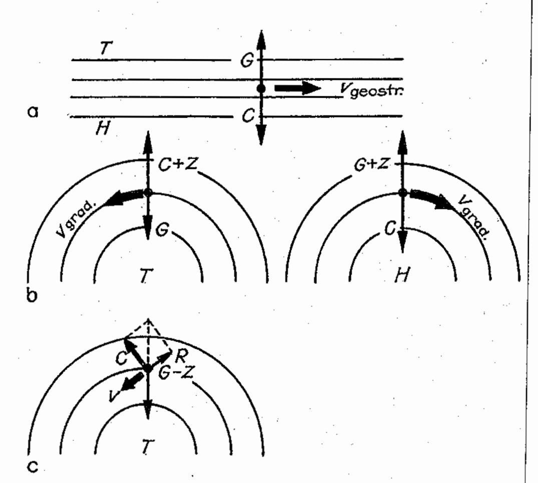


Abb. 10.4.2–4 Schematische Darstellung (a) des geostrophischen Windes, (b) des Gradientwindes und (c) des Windes unter Einfluß der Reibung

es erfolgen geringfügige, meist innerhalb der Meßgenauigkeit liegende Abweichungen vom Gradientwind, die zu einem entsprechenden Massenfluß und damit zur Veränderung des Druckfeldes führen. Die in der Strömung horizontal transportierten Luftmassen (*Advektion*) unterschiedlicher Temperatur und Dichte führen zu spezifischen Wettererscheinungen und wirken auch auf die Veränderung des Druckfeldes ein.

In der unteren Atmosphäre ist die Reibung der Luft an der Erdoberfläche nicht vernachlässigbar klein. Es entsteht ein Windfeld nach dem Kräfteschema der Abb. 10.4.2–4c, die Windrichtung weicht mit einer Komponente zum tieferen Druck vom Gradientwind ab und die Geschwindigkeit ist geringer. Dieser Reibungseinfluß ist abhängig von der Rauhigkeit der Unterlage und der Schichtungsstabilität. Er wird mit zunehmender Höhe geringer, so daß der wahre Wind schließlich mit dem Gradientwind übereinstimmt (das ist im Mittel ab 800 m über Grund). Diese von der Reibung unmittelbar beeinflußte untere Atmosphärenschicht ist die *planetarische Grenzschicht*; über ihr liegt die *freie Atmosphäre*.

Für viele Probleme, z. B. die Ausbreitung von Emissionen (vgl. 10.5.2.2.) sind die speziellen Eigenschaften und Gesetzmäßigkeiten der Windstruktur der atmosphärischen Bodenschicht bedeutsam. Das vertikale Windprofil in Bodennähe wird durch die *Schubspannung* $\tau = A(\partial v/\partial z)$ gekennzeichnet, wobei der Austauschkoeffizient A von der Höhe und der Schichtungsstabilität abhängt. Nach experimentellen Erfahrungen ist in den unteren 2 bis 3 Dekametern τ = const. mit der Höhe. Unter Verwendung der *Schubspannungsgeschwindigkeit* $u_x = \sqrt{(\tau/\varrho)}$ ergibt sich hier für das mittlere vertikale *Windprofil* $[v(z)]\ \bar{v}(z) = u_x k^{-1} \ln(z/z_0)$, mit $k = 0{,}4$ = von Kármán-Konstante und z_0 = Rauhigkeitsparameter = Höhe, in der $v(z) = 0$ wird (z_0 = für Schneeflächen 0,01 bis 0,1 cm, für Wiese 0,1 bis 10 cm, für Wald und Großstadt 50 bis 300 cm).

Über dieser „Prandtl-Schicht" mit $\tau(z)$ = const. erstreckt sich bis zur freien Atmosphäre (mit $v = v_{grad}$) die „Ekman-Schicht", in der das vertikale Windprofil durch ein Potenzgesetz angenähert werden kann: $v_2/v_1 = (z_2/z_1)^m$. Charakteristische Werte für m sind 0,5 für stabile, 0,15 für indifferente und 0,10 für labile Schichtung.

10.4.3. Allgemeine Zirkulation, Wetter und Klima

10.4.3.1. Allgemeine Zirkulation. Unter *allgemeiner Zirkulation* ist der charakteristische mittlere atmosphärische Strömungsverlauf mit periodischen und unregelmäßigen Veränderlichkeiten zu verstehen. Der Antrieb der allgemeinen Zirkulation liegt in der unterschiedlichen Strahlungsbilanz, die neben der Breite von der Art der Erdoberfläche u. a. Parametern abhängt. Zusammen mit den Meeresströmungen bewirkt sie einen Wärmetransport und einen teilweisen Temperaturausgleich, wodurch sich ein charakteristischer Zustand herausbildet, der die Klimazonen der Erde bestimmt.

Die allgemeine Zirkulation wird meist in Form mittlerer Druckfelder einzelner Niveaus dargestellt. Hauptglieder der allgemeinen Zirkulation sind die subtropischen Hochdruckzellen mit der dazwischenliegenden innertropischen Konvergenzzone, die zyklonalen Polarwirbel und die Westwindzonen der mittleren Breiten. Die polaren Tiefdruckwirbel erstrecken sich im Winter auch durch die Stratosphäre, während im Sommer (durch Absorption der Sonnenstrahlung, besonders des UV durch das Ozon) sich dort ein stratosphärisches Hoch ausbildet.

Die Hochdruckzonen sind wegen der stabilisierenden Wirkung großräumig absinkender Luftbewegungen ein Trokkengebiet. Der *Passat*, die beständigste irdische Luftströmung transportiert wasserdampf- und wärmereiche Luft der unteren Atmosphäre in niedere Breiten, wo die angereicherte Energie und Feuchte freigesetzt wird. *Monsune* sind mit der Verschiebung der Glieder der allgemeinen Zirkulation im Jahresgang periodisch wechselnde Zirkulationsformen und damit mit der allgemeinen Zirkulation in der Troposphäre zusammenhängende Wetter- und Witterungserscheinungen. Besonders ausgeprägt ist der Monsum Indiens, der einen Wechsel zwischen trockenem Nordostpassat (Winter) und dem Einfluß der tropischen Tiefdruckzone (Sommer) darstellt.

In die *Westwindzone* der gemäßigten Breiten, der Übergangszone zwischen subtropischen Warmluft- und polaren Kaltluftmassen, sind in Form meist rasch wandernder Hoch- und Tiefdruckwirbel „Makroturbulenzkörper" eingelagert, die neben den Meeresströmungen den horizontalen Energieaustausch zwischen den Zonen mit Überschuß bzw. Defizit der Strahlungsbilanz regulieren und den wechselhaften Charakter von Wetter und Witterung in den mittleren Breiten verursachen.

10.4.3.2. Wetter. *Wetter*, *Witterung* und *Klima* sind komplexe Erscheinungsformen eines außerordentlich vielfältig verzweigten Systems dynamischer und thermodynamischer Prozesse in dem nach außen offenen System Atmosphäre/Erdoberfläche in verschiedenen Zeitmaßstäben. Die wetterbildenden Prozesse sind der Hauptgegenstand der *Meteorologie*, speziell der *„synoptischen Meteorologie"* (synoptisch = zusammenschauend). Die Komplexität dieser Prozesse zwingt zur Arbeit mit vereinfachten Modellen, die die wesentlichen Prozesse enthalten und die für Vorausbestimmungen von Wetterentwicklungen und -erscheinungen brauchbar sind.

Ein solches anschaulich vereinfachtes Modell beinhaltet die *Idealzyklone*, die dominierenden Prozesse und Merkmale innerhalb einer Zyklone der mittleren Breiten (Abb. 10.4.3–1). Diese wird als Entwicklungsprozeß von einer Wellenstörung an der *Polarfront*, der Übergangszone zwischen Warmluft- und Kaltluftmassen, bis zu einem voll ausgebildeten kreisförmigen Tiefdruckfeld aufgefaßt. Durch Vordringen der Warmluft gegen die vorgelagerte Kaltluft entsteht als Übergangszone mit starkem Temperaturgradienten die *Warmfront*, während die rückseitige Kaltluft in aktivem Vordringen gegen die Warmluft eine *Kaltfront* bildet. Die Warmfront ist gekennzeichnet durch Gleitbewegungen der Warmluft auf die Kaltluft mit Hebung, adiabatischer Abkühlung und (Schicht-)Wolkenbildung, meist mit Niederschlag. Die Kaltfront existiert in der Regel wegen der Windzunahme mit der Höhe als eine sehr steile Übergangszone, die den Charakter einer labilisierten Umlagerungszone hat mit entsprechender Quellbewölkung, mit Schauern und Gewittern. Häufig, bei bereits alternden Zyklonen, ist der zwischen den Wetterfronten liegende *Warmsektor* soweit eingeengt, daß die Rückseitenkaltluft auf die Vorderseitenkaltluft aufläuft; dieses Stadium einer *Okklusion* erscheint als kontinuierlicher Übergang der Warmfront- in Kaltfronterscheinungen oder überhaupt nur als Kaltfronttyp.

Im Einflußbereich von *Hochdruckgebieten* herrscht wegen der ageostrophischen Massendivergenz in der planetarischen Grenzschicht großräumige Absinkbewegung (in der

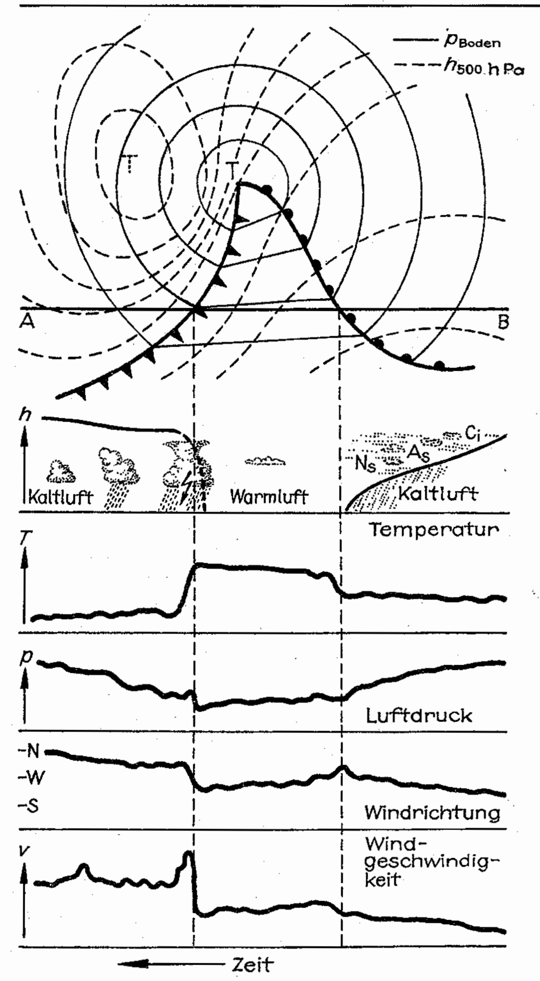


Abb. 10.4.3-1 Modell der Idealzyklone

Größenordnung mm/s), die zur Stabilisierung der Schichtung und zur Austrocknung in der freien Atmosphäre, damit zu Wolkenarmut und hoher Ein- und Ausstrahlung führen mit entsprechend relativ großen täglichen Temperaturamplituden. Im Winter bildet sich in der planetarischen Grenzschicht oft Nebel oder Hochnebel aus.

Für die Vorausbestimmung werden, ausgehend von einer gründlichen Analyse des Ausgangszustandes, in die Messungen und Beobachtungen vom Boden, aus der freien Atmosphäre und von Satelliten einfließen, empirische, grafische, statistische und numerische Methoden gebraucht. Die gegenwärtige Vorhersagepraxis arbeitet mit mehrschichtigen Modellen nach unterschiedlichen Methoden, mit denen unter schrittweiser Integration kinematischer oder dynamischer Parameter eines hinreichend dichten Netzes von Gitterpunkten relativ sichere Vorausbestimmungen von Druckfeldern der freien Atmosphäre möglich sind. Ein besonders schwieriges Problem ist die Umsetzung solcher Druckfeldkarten in das zugehörige Wetter.

10.4.3.3. Klima. Das *Klima*, der mittlere Zustand der Atmosphäre und seiner charakteristischen Schwankungsbreite über einem bestimmten Ort und bezogen auf eine bestimmte Zeitepoche, wird nach speziell definierten Klassifikationen („effektiven“, die die Wirkung beschreiben, und „genetischen“, die von den Ursachen ausgehen) in Form von Verteilungen der *Klimaelemente* dargestellt. Es wird durch eine große Zahl terrestrischer und extraterrestrischer Einflüsse, durch die *Klimafaktoren* bestimmt. Im Prinzip ist das Klima die Erscheinungsform des Wärme- und Wasserhaushaltes. Die wesentlichen bestimmenden Klimafaktoren sind

- die extraterrestrische Intensität der Sonnenstrahlung und ihre spektrale Verteilung,
- die Intensität und spektrale Verteilung der Wärmestrahlung der Erdoberfläche und der Atmosphäre,
- die Albedo der Erdoberfläche und des Gesamtsystems Erdoberfläche/Atmosphäre,
- die atmosphärische Trübung (in Abhängigkeit von der Wellenlänge), orts- und zeitabhängiger Gesamtgehalt einer Luftsäule an Wasserdampf, Kohlendioxid, Ozon und die Bewölkung,
- die Oberflächentemperatur der Ozeane und
- die Intensität, Lage und Eigenschaften der Hauptströmungen der Atmosphäre und der Ozeane.

Die Suche nach Erklärungen von *Klimaänderungen* in der Erdgeschichte und nach Vorausbestimmung zukünftiger Veränderungen ist gleichbedeutend der Frage nach Veränderungen der Klimafaktoren und nach dem Wechselwirkungsmechanismus, der bei Veränderung eines Faktors ausgelöst wird. Gegenwärtig existierende Vorstellungen über solche Mechanismen sind noch weitgehend hypothetisch. Grundsätzlich sind sowohl terrestrische als auch extraterrestrische Faktoren als realistische Ursachen nachgewiesener Änderungen möglich. Auch anthropogene Eingriffe (vgl. 10.5.2.) durch Wärmezufuhr in die Atmosphäre und durch Veränderungen der spektralen Transparenz müssen als potentiell verändernde Faktoren angesehen und entsprechend untersucht und überwacht werden.

10.4.4. Luftelektrizität

Die weltweite *Gewitter*tätigkeit realisiert in der Art eines Dynamos die Beständigkeit eines globalen luftelektrischen Zustandes. Bei ungestörtem „Schönwetter“ existiert in Bodennähe ein vertikales Spannungsgefälle von 130 V/m, das in 10 km Höhe nur noch etwa 4 V/m beträgt.

In grober Näherung ist der luftelektrische Zustand dem eines Kugelkondensators vergleichbar, zwischen dessen leitenden Flächen (Erdoberfläche und Ionosphäre) mit einer *Potentialdifferenz* von 200 bis 250 kV das Dielektrikum Luft liegt. Die (geringe) *Leitfähigkeit* der Luft beruht auf der Beweglichkeit der Luftionen; sie liegt bei $\approx 2{,}5 \cdot 10^{-16}\,\Omega^{-1}\mathrm{cm}^{-1}$. Die *Vertikalstromdichte* beträgt in der ungestörten Atmosphäre $\approx 3 \cdot 10^{-12}\,\mathrm{A/m^2}$.

Die in jedem Augenblick wirkenden $\approx 2\,000$ Gewitter (vor allem im tropischen Bereich) müssen zur Aufrechterhaltung des luftelektrischen Feldes so wirken, daß je Gewitter ≈ 1 A fließt und in der Höhe positive Ladung erzeugt wird. Für den Mechanismus der Ladungstrennung in der Schauer- und Gewitterwolke gibt es noch keine eindeutige Erklärung, obwohl zahlreiche „Gewittertheorien“ von physikalisch realistischen Wirkprozessen ausgehen, meist aber nur Teilprozesse beschreiben.

Der *Blitz* ist die Erscheinung eines Raumladungsausglei-

ches in einem Blitzkanal, der durch Stoßionisation an Orten mit hohem Potentialgefälle (2 bis 3 kV/m) elektrisch leitend wird (in Form des beobachtbaren ruckweisen Vordringens des Blitzkanals), bis eine Verbindung mit der Erdoberfläche oder mit einem anderspoligen Ladungszentrum einer Wolke hergestellt ist. Mit einer Zeitdauer von $\approx 4 \cdot 10^{-5}$ s erfolgt als Blitz dann der Ladungsausgleich, dem häufig mit Abständen von $\approx 3 \cdot 10^{-2}$ s weitere Entladungen folgen. Die mittlere Stromstärke eines Blitzes liegt bei 10^4 bis 10^5 A. Wegen der kurzen Entladungszeit wird aber nur eine Energie von wenigen 10 kWh transportiert. Expansion und folgendes Zusammenschlagen des erhitzten Blitzkanals wird akustisch als *Donner* wahrgenommen.

10.4.5. Physik der Hochatmosphäre

Als *Hochatmosphäre* bezeichnet man den oberhalb des stratosphärischen Windminimums bei 24 km Höhe und damit auch oberhalb des Maximums der Ozonkonzentration liegenden äußeren Teil der Atmosphäre bis zur Exosphäre. Sie unterscheidet sich physikalisch von der unteren Atmosphäre durch ihre vorzugsweise Strahlungsabsorption im gesamten UV-Bereich mit sehr eigenständigen Energieumsetzungen, besonders fotochemische Prozesse, Dissoziation und Ionisation.

Die Absorption im solaren UV erfolgt vor allem durch Sauerstoff O_2 ($\lambda < 220$ nm) und Ozon O_3 ($\lambda = 195$ bis 320 nm). Atomarer Sauerstoff O und Stickstoff N_2 absorbieren zusätzlich im Bereich $\lambda < 100$ nm. Angesichts des bei Wellenlängen über 200 nm stark zunehmenden solaren Strahlungsflusses kommt dem O_3 die maßgebliche Rolle für das Temperaturregime der Hochatmosphäre zu. Strahlungsbilanzrechnungen ergeben, daß das beobachtete Temperaturfeld nicht aus einem Strahlungsgleichgewicht erklärt werden kann, sondern daß dynamische Prozesse, wahrscheinlich auch Energieaustausch zwischen beiden Halbkugeln, existieren müssen.

Untersuchungen der Windsysteme, die unter Verwendung von Raketen und durch *Driftmessungen* mit Radar an ionisierten Meteoritenschweifen oder mit Hilfe der Irregularitäten der Elektronenkonzentrationen der Ionosphäre, die sich in Radiowellenfadings widerspiegeln, durchgeführt werden, ergeben folgende wesentlich Merkmale: Vorherrschend gleichmäßige zonale Ostwinde ($\triangleq$ Driftrichtung nach West) mit maximal 60 m/s im Höhenbereich von 65 bis 70 km im Sommer werden im Winter durch wesentlich stärkere (Geschwindigkeit > 100 m/s im Niveau 50 km), in entgegengesetzter Richtung strömende und stark veränderliche Winde ersetzt.

Diese Zirkulation des stratosphärisch-mesosphärischen Bereiches wird durch jahreszeitlich veränderliche Temperaturverteilung in ähnlicher Weise wie die troposphärische angetrieben, jedoch stärker durch Gezeitenschwingungen überlagert. Mit zunehmender Höhe stellt schließlich die thermosphärische Zirkulation ein System dar, dessen Ursprung in der tagesperiodischen thermischen Expansion und Kontraktion liegt.

Als *Ionosphäre* werden die Schichten bezeichnet, in denen hoher Ionisationsgrad auftritt. Dieser wird durch die sehr kurzwellige Sonnenstrahlung verursacht (die Ionisationskanten für die atmosphärischen Hauptbestandteile liegen für Stickstoff N_2 bei 79,6 nm, für Sauerstoff O_2 bei 102,7 nm und für atomaren Sauerstoff O bei 91,0 nm), die hierfür praktisch vollständig in der Thermosphäre aufgebraucht wird. Wegen der Schichtungsstruktur nach den Bestandteilen in dieser „Heteropshäre" (vgl. 10.1.4.3.) entstehen höhenabhängige Ionisationsmaxima: Das Hauptmaximum im Höhenbereich 250 bis 350 km, die *F-Schicht*, mit charakteristischen Elektronendichten um $10^6\,cm^{-3}$ und vorwiegend O^+-Ionen, darunter, mit Maximum bei 100 bis 120 km, das sekundäre Maximum, die *E-Schicht* mit Elektronendichten um $10^5\,cm^{-3}$ und vorwiegend O_2^+-Ionen und NO^+-Ionen, die aus primären N_2^+-Ionen entstanden sind. Im oberen Mesosphärenbereich liegt die *D-Schicht* (Elektronendichten $< 10^4\,cm^{-3}$), die sich wahrscheinlich aus der Dissoziation des Spurengases Stickstoffmonoxid NO (durch $\lambda \leqq 134$ nm) bildet.

Dichte und Höhenverteilung der freien Elektronen unterliegen Tages- und jahresperiodischen Schwankungen und aperiodischen Veränderungen durch die Sonnenaktivität (vgl. 10.1.5.). Der Ionosphärenzustand hat großen Einfluß auf den Funkverkehr. Im Lang- und Mittelwellengebiet erfolgt tagsüber eine Dämpfung im unteren Teil der Ionosphäre, nachts eine Erhöhung der Reichweite. Kurzwellenausbreitung erfolgt nur über ionosphärische Reflexion in Abhängigkeit von der Elektronendichte und der Wellenlänge. Ultrakurzwellen sind solche Wellenlängen, die im Normalfall an der Ionosphäre nicht mehr reflektiert werden, also oberhalb einer *Grenzfrequenz* liegen.

10.5. Umweltgeophysik

Der aus der Ökologie stammende Begriff *Umwelt* bedeutet die Gesamtheit aller auf die Lebensvorgänge (von Pflanze, Tier und Mensch) einwirkenden äußeren Zustände und Prozesse. Gegenstand der *Umweltgeophysik* sind die Wechselbeziehungen der Biosysteme (besonders des Menschen) mit den physikalischen Bedingungen auf dem Planeten Erde.

10.5.1. Natürliche geophysikalische Umweltfaktoren

Die Entwicklung und Existenz von Lebewesen auf der Erde (vgl. 10.1.2.) vollzog und vollzieht sich unter dem Einfluß physikalischer Zustände und Felder mit teils periodischen oder aperiodischen Veränderlichkeiten. Zu den wesentlichen *geophysikalischen Umweltfaktoren* zählen:

- die Erdschwere,
- das Erdmagnetfeld,
- die Erdrotation mit dem Tag-Nacht-Rhythmus,
- die Energie der Sonnenstrahlung, die zusammen mit dem modifizierenden Prozessen das Klima bildet,
- die physiologische Wirkung spezieller (besonders auch der ionisierenden) Strahlungsformen und -frequenzbereiche.

Die *Biogeophysik* untersucht Formen und Mechanismen der Einwirkung natürlicher geophysikalischer Zustände und Prozesse auf Pflanze, Tier und Mensch; als Teilgebiet wird besonders intensiv auf dem Gebiet der Einwirkung von Klima und Wetter auf die Organismen (*Bioklimatologie* oder *Biometeorologie*) geforscht. Kopplungen zwischen geophysikalischen Bedingungen und physiologischen Reaktionen können in sehr zahlreichen Formen statistisch nachgewiesen, in ihren Wirkungsmechanismen aber meist noch ungenügend erklärt werden. Besonders ist die Kopplung physiologischer Rhythmen mit entsprechenden geophysikalischen ein interessantes Arbeitsgebiet geworden.

10.5.2. Anthropogene Eingriffe in natürliche geophysikalische Umweltfaktoren

Auch die biologische Evolution ist ein Prozeß mit Wechsel- und Rückwirkungen auf den geochemischen und den geophysikalischen Umweltzustand (vgl. 10.1.2.). Die Tätigkeit des Menschen greift mit seiner Entwicklung regional, mit zunehmender Technisierung auch global in die natürlichen Bedingungen ein. Kontrolle, Schutz und auch Melioration der natürlichen Umwelt sind internationale wissenschaftlich-technische Aufgaben geworden.

che (≈ 160 W/m²). Weitere Zunahme der Energieproduktion (gegenwärtig 8 %/Jahr) macht diesen Eingriff für den natürlichen Wärmehaushalt nicht mehr vernachlässigbar. Über Großstädten und Industriegebieten liegt die Wärmeabgabe im Bereich von 10^2 bis 10^3 W/m² und trägt neben den Veränderungen des natürlichen Wärmehaushaltes durch die Bebauung zu den regionalen Klimaanomalien dieser Gebiete bei, die sich u. a. in einer Erhöhung der Temperatur gegenüber der Umgebung äußern.

10.5.2.2. Eingriffe über Verunreinigungen der Geosphären. Ausbreitung anthropogener Emissionen. Die *anthropogenen Emissionen* (besonders durch Verbrennung fossiler Energieträger, durch chemische Industrien und durch den Verkehr) in gasförmiger, flüssiger und fester Form breiten sich im Boden, im Wasser oder in der Luft aus. Im Boden dringen Kontaminationen durch *Migration* in den Kapillaren und durch Bindung an das Bodenwasser in tiefere Schichten, teils in das Grundwasser ein. In stehenden oder fließenden Gewässern werden Beimengungen durch die *turbulente Diffusion* (so bezeichnet wegen vieler Ähnlichkeiten mit dem Vorgang der molekularen Diffusion) auf benachbarte Schichten oder Wasserkörper übertragen, dabei verdünnt, als Sedimente am Boden abgelagert oder mit dem fließenden Wasser bis in die Meere transportiert.

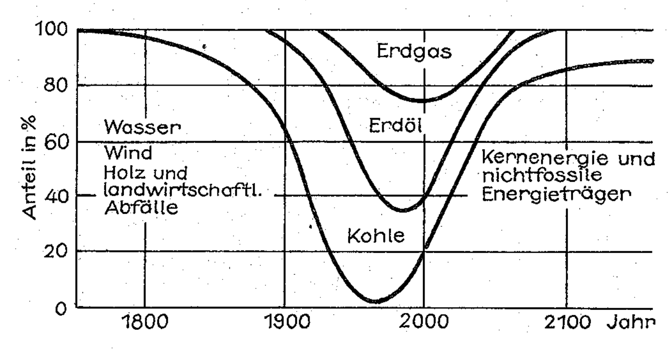


Abb. 10.5.2–2 Entwicklung des Anteiles verschiedener Primärenergieträger an der Energiegewinnung (nach Alexander)

10.5.2.1. Direkte Eingriffe erfolgen besonders durch ionisierende Strahlungen. Die Tab. 10.5.2–1 zeigt durchschnittliche *Strahlenbelastungen*, denen der Mensch in entwickelten Industriestaaten ausgesetzt ist. Aus gegenwärtiger Sicht wird sich die Entwicklung der Nutzung von Primärenergieträgern nach Abb. 10.5.2–2 vollziehen. Von Kernkraftwerken werden auch im Normalbetrieb radioaktive Stoffe (besonders Tritium) in die Gewässer abgegeben; wie Erfahrungen zeigen, ist aber die radioaktive Kontamination der Flüsse durch Kernkraftwerke kleiner als die natürliche. Die in die Atmosphäre eindringenden Radionuklide bestehen im wesentlichen aus Edelgasen (^{133}Xe) und aktivierten Gasen (^{131}I, ^{14}Ar, ^{16}N) und liegen um den Faktor 10^{-3} unter zulässigen Grenzwerten. Anthropogene Quellen radioaktiver Belastung sind vor allem die Experimente mit nuklearen Sprengsätzen.

Die anthropogene Energieproduktion der vielfältigen Formen führt zu direkter Wärmeabgabe an die Atmosphäre. Bei der derzeitigen *Weltenergieproduktion* in der Größenordnung 10^{13} W erreicht die über die Erdoberfläche gemittelte Leistungsdichte ≈ 0,02 W/m², beträgt somit das 10^{-4}fache der mittleren Solarstrahlungsabsorption an der Erdoberflä-

Tab. 10.5.2–1 Durchschnittliche jährliche Strahlenbelastung eines Menschen unter mitteleuropäischen Verhältnissen

Strahlungsquelle	Strahlenbelastung in mSv
kosmische Strahlung	0,50
natürliche radioaktive Stoffe	0,60
medizinische Strahlenanwendung	0,50
natürliche Radioaktivität in Nahrung und Atemluft	0,25
Kernexplosionen	0,08
Kernkraftwerke	< 0,01

Wegen der hohen Beweglichkeit der Luft sind die Transportprozesse in der Atmosphäre besonders vielfältig und kompliziert. Die *atmosphärische Turbulenz* verteilt die Beimengungen auf ein größeres Luftvolumen und verringert dabei die spezifische Konzentration (*Immission*). Diese atmosphärische Turbulenz setzt sich aus der dynamisch (durch Rauhigkeit am Boden oder an Schichtgrenzen) und thermisch (durch Schichtungsinstabilität) bedingten Turbulenz zusammen und ist eine von den meteorologischen Bedingungen abhänigige sehr variable Größe.

Mit der großräumigen atmosphärischen Zirkulation werden die Beimengungen horizontal transportiert. Auf diesem Wege unterliegen sie physikalisch-chemischen Stoffumwandlungen. Durch *Ausfallen* und *Auswaschen* mit dem Niederschlag gelangen die Stoffe wieder zur Erdoberfläche. Mit numerischen *Ausbreitungsmodellen* (vereinfachten, auf die wesentlichen handhabbaren Parameter reduzierten Gleichungen der Transportprozesse) wird versucht, die Stofftransporte und die damit zusammenhängenden Immissionen aufgrund bekannter Quellstärken für unterschiedliche Zeit- und Raumskalen näherungsweise zu bestimmen.

Veränderungen geophysikalischer Umweltzustände durch atmosphärische Immissionen. Neben den hygienischen und geochemischen Konsequenzen wirken *Luftbeimengungen* verändernd auf Komponenten des Strahlungshaushaltes

(vgl. 10.4.1.2.) ein. Die regional hohen Luftverunreinigungen über Großstädten und Industriegebieten vergrößern über Absorption und Rückstreuung die *Extinktion* (vgl. 10.4.1.1.) so, daß hier die Globalstrahlung gegenüber dem umgebenden freien Land um ≈ 10 % geringer sein kann. Im kurzwelligen (Solar-)Strahlungsbereich wirkt besonders das *Aerosol*, das bei hohen Luftfeuchten durch Anlagerung an Wasser starken *Dunst* bildet.

Gasförmige anthropogene *Luftbeimengungen* absorbieren vor allem im Bereich der terrestrischen Ausstrahlung mit der Tendenz, den atmosphärischen Glashauseffekt zu verstärken. Grundsätzlich besteht die Möglichkeit, daß anthropogene Luftbeimengungen über Veränderungen des Strahlungshaushaltes in den Wärmehaushalt und damit in den komplexen Umweltfaktor *Klima* verändernd eingreifen. Das gilt besonders für das atmosphärische CO_2, das infolge ständig zunehmender Verbrennung fossiler Energieträger (Kohle, Öl, Gas) trotz der Wirkung starker Senken (Landvegetation, Ozeane) im globalen Maßstab von ≈ 290 ppm (im Jahre 1880) auf ≈ 330 ppm (im Jahre 1980) angestiegen ist. Die mit unterschiedlichen Modellen abgeschätzte erhöhte Glashauswirkung, ausgedrückt in Erhöhung der Mitteltemperatur an der Erdoberfläche, bei Erhöhung der Immissionskonzentration einiger wichtiger Gase ist aus Tab. 10.5.2–3 ersichtlich.

Die Wirkung des festen *Aerosols* auf den Strahlungshaushalt ist neben der Immissionsverteilung stark von den optischen Eigenschaften der Teilchenkollektive abhängig, die wiederum mit der Luftfeuchte veränderlich sind. Der Refraktionsindex des komplexen realen Aerosols beträgt im Durchschnitt im Realteil $m_r = 1{,}44$ bis 1,69, im Imaginärteil, der die Absorptionseigenschaften darstellt, variiert er zwischen 10^{-1} und 10^{-3}. Über einer Erdoberfläche hoher Albedo setzt erhöhte atmosphärische Partikelimmission die planetare Albedo herab, das System Erdoberfläche/Atmosphäre gewinnt dort Energie aus der Sonnenstrahlung. Bei kleiner Erdoberflächenalbedo (Ozeane!) ist das umgekehrt wegen der durch das Aerosol vergrößerten planetaren Albedo infolge verstärkter Rückstreuung. Der thermische Effekt veränderter Partikelimmission wird stark durch das Absorptionsvermögen des Aerosols mitbestimmt, sowohl für kurzwellige wie für langwellige Strahlung.

Eis- und Wolkenbedeckung des Planeten Erde sind sehr maßgebliche Einflußgrößen im Strahlungshaushalt; Abschätzungen anthropogener Klimaänderungen müssen deshalb Aussagen über damit verbundene Veränderungen dieser Größen einschließen. Dazu bestehen erst unzureichende Ansätze. Die Temperaturschwankungen der vergangenen 100 Jahre enthalten keine signifikante Information über anthropogene Veränderungen. Eine gute Korrelation zwischen vulkanischer Aktivität und resultierender Trübungsschwankung mit dem Temperaturgang lassen vermuten, daß hierin die Ursache für Mitteltemperaturschwankungen liegt. Sowohl die Steuerung ungewollter, nachteiliger anthropogener Klimaänderungen wie auch eine gezielte großräumige Klimaoptimierung sind Aufgaben der Zukunft, die eine genauere Kenntnis der Klimabildung voraussetzen. Arbeiten an einem hinreichend vollständigen numerischen *Klimamodell* sind deshalb ein Schwerpunkt geophysikalischer Forschung.

Tab. 10.5.2–3 Die Glashauswirkung atmosphärischer Spurenstoffe

Stoff	Schwerpunkt der Absorptionsbanden in nm	gegenwärtige atmosphärische Konzentration in ppm	errechnete Auswirkung an der globalen Mitteltemperatur bei Konzentrations-Vervielfachung (a) a	in K
NO_2	7 780	0,28	2	0,68...0,44
	17 000			
	4 500			
CH_4	7 660	1,6	2	0,28...0,20
NH_3	10 530	$6 \cdot 10^{-3}$	2	0,12...0,09
HNO_3	5 900	10^{-4} (?)	2	0,08...0,06
	7 500			
	11 300			
	21 800			
C_2H_4	10 500	$2 \cdot 10^{-4}$	2	0,01...0,001
SO_2	8 690	$2 \cdot 10^{-3}$	2	0,03...0,02
	7 350			
CCl_2F_2	9 130	$1 \cdot 10^{-4}$	20	}
	8 680			0,54...0,36
	10 930			}
CCl_3F	9 220	$1 \cdot 10^{-4}$	20	}
	11 820			
CH_3Cl	13 660	$5 \cdot 10^{-4}$	2	}
	9 850			0,02...0,01
	7 140			}
CCl_4	12 990	$1 \cdot 10^{-4}$	2	}
CO_2	15 000	330	2	1,5...2,5

Nachdem die katalytische Wirkung der Chlorfluormethane auf den Abbau von Ozon (O_3) erkannt ist, ist auch die grundsätzliche Möglichkeit einer anthropogenen Zerstörung der stratosphärischen Ozonosphäre in der Diskussion. Eine Verringerung des im UV unterhalb 290 nm Wellenlänge stark absorbierenden Ozons würde die Transparenz für diese biologisch schädliche Strahlung erhöhen und an der Erdoberfläche physiologische Konsequenzen haben. Auch hier liegt gegenwärtig noch kein Nachweis aus der Natur vor.

Wegen der sehr engen Verflechtungen chemischer und physikalischer Prozesse und Zustände ist die geophysikalische Umweltproblematik sehr unmittelbar mit der geochemischen verbunden.

11. Biophysik

11.1. Wesen und Inhalt der Biophysik

Es ist bei der Biophysik – wie auch im Falle anderer Grenzgebiete – sehr schwer, wenn nicht gar unmöglich, scharfe Trennlinien zu benachbarten Disziplinen (hier u.a. Biochemie, Molekularbiologie, Physiologie) zu ziehen, zumal sich gerade auch die Biophysik in einer äußerst dynamischen Entwicklung befindet. Auf keinen Fall ist der Begriff Biophysik so auszulegen, als wäre diese eine „besondere" Physik der lebenden Materie mit eigenen spezifischen Grundgesetzen, die sich von der „gewöhnlichen" Physik in irgendeiner Weise unterscheiden würde. Es gibt keinen Grund zu der Annahme, daß für die Beschreibung und das Verständnis der biologischen Phänomene andere als die aus Physik und Chemie bekannten Grundgesetze notwendig wären.
Je nachdem, unter welchem Blickwinkel die Biophysik gesehen wird bzw. welche Aspekte man für entscheidend hält, kann man sie sowohl als Teilgebiet der Biowissenschaften wie auch als einen Zweig der Physik einordnen. Entsprechend einer verbreiteten Auffassung kann unter Biophysik die Anwendung und Übertragung von in der Physik entwikkelten und bewährten Denkweisen und Methoden auf die Analyse von Struktur und Funktion biologischer Systeme verstanden werden (GLASER).
Die biophysikalische Analyse durchdringt in zunehmendem Maße alle Bereiche der biowissenschaftlichen Forschung. Da es im Rahmen der vorliegenden Darstellung aussichtslos erscheint, allen angeführten Problemen und Inhalten der modernen biophysikalischen Forschung auch nur annähernd Rechnung zu tragen, sind im folgenden einige Schwerpunktprobleme ausgewählt und erläutert, die sowohl eine bedeutende Rolle in der aktuellen Forschung spielen als auch die Ziele, Inhalte und Methoden der modernen Biophysik erkennen lassen und überdies von allgemeinem Interesse sein dürften. Hierbei müssen zwangsläufig zahlreiche wichtige Richtungen, insbesondere der „klassischen" Biophysik (z.B. Biomechanik, wie die Physik des Schwimmens, des Fliegens, des Blutkreislaufs und des Atemgastransports, oder die Populationsdynamik u.a.) leider zu kurz kommen, was aber keinerlei Wertung ihrer Bedeutung beinhalten soll.
In vielen der nachstehend beschriebenen Gebiete und Forschungsergebnisse der Biophysik steckt die „eigentliche" Physik in den (oft nur am Rande erwähnten) physikalischen Methoden, mit denen diese Resultate erzielt werden konnten. Häufig wurden diese Ergebnisse erst durch Weiterentwicklung, Modifizierung und Anpassung bekannter physikalischer Meßprinzipien und -verfahren an die spezielle biophysikalische Problematik möglich.

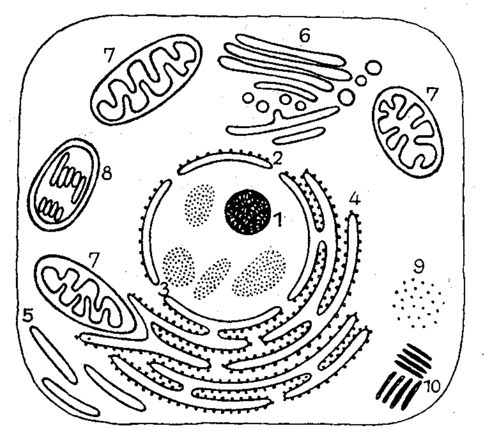


Abb. 11.2.1–1 Schematische Darstellung einer eukaryotischen Zelle (nicht maßstabgerecht!); 1 Nukleolus, 2 Kernhülle (Doppelmembran), 3 Kernpore, 4 rauhes endoplasmatisches Retikulum, 5 glattes endoplasmatisches Retikulum, 6 Golgi-Apparat, 7 Mitochondrien, 8 Chloroplast, 9 Ribosomen, 10 Zentriole (Zentrosom)

11.2. Zelluläre und molekulare Basis der Lebensfunktionen

11.2.1. Zelle und Zellorganellen

11.2.1.1. Die Zelle als Grundeinheit des Lebens. Die Zelle stellt die kleinste Einheit des Lebendigen dar, die alle das Leben kennzeichnenden Prozesse, nämlich Stoffwechsel, Selbstreproduktion, Mutabilität, Erregbarkeit, Wachstum als die wichtigsten, noch koordiniert durchführen kann. Trotz der Vielzahl verschiedener Zelltypen sind alle Zellen nach ähnlichen Prinzipien aufgebaut, d.h. sie zeigen eine relativ einheitliche Organisation und innere Struktur. Dabei unterscheidet man zwischen den Zellen der sogenannten *Prokaryoten* (*Bakterien*, *Cyanobakterien* oder Blaualgen), den *Protozyten*, d.h. Zellen ohne echten Zellkern, und den Zellen der *Eukaryoten (Euzyten)* mit einem echten Zellkern (alle Vielzeller, aber auch einige Einzeller).
Thermodynamisch stellt die Zelle ein *offenes System* dar, in welchem ständig eine Unzahl verschiedener Reaktionen abläuft – Auf- und Abbau von Nahrungsstoffen, „Baumaterialien", usw., ohne daß sich in der hohen räumlichen und zeitlichen Ordnung dieser Prozesse ein „Nachlassen" bemerkbar machen würde – im Gegenteil, bei einer Änderung der Umgebungsbedingungen kann die Zelle auch hierauf sinnvoll reagieren. Die Erhaltung (bzw. sogar Vergrößerung) dieser Ordnung in einer viel weniger komplexen (geordneten) Umwelt ist nur möglich durch die ständige Aufnahme und Verwertung *(Dissipation)* von Energie. Dadurch wird die Zelle in einem als *„Fließgleichgewicht"* (engl. *steady state*) bezeichneten Zustand weitab vom *thermodynamischen Gleichgewicht* gehalten, der durch stete Ein- und Ausfuhr von Synthese- bzw. Abbauprodukten und eben Energieaustausch gekennzeichnet ist.
Eine wichtige Voraussetzung für das geordnete Nebeneinander der verschiedenen biochemischen Reaktionen, insbe-

sondere in den eukaryotischen Zellen, bildet die Unterteilung („*Kompartimentierung*") der Zelle in unterschiedliche Reaktionsräume, die gegen das *Grundplasma* (auch *nukleozytoplasmatische Matrix*) durch Membransysteme abgeteilt sind und ganz spezifische Aufgaben im Zellstoffwechsel zu erfüllen haben, z. B. Energiebereitstellung.

11.2.1.2. Größe und Bau von Zellen. Im allgemeinen sind Zellen selten wesentlich größer als ≈100 µm bzw. kleiner als ≈1 µm im Durchmesser. Bekannteste Ausnahmen für größere (Körper-)Zellen (neben den Eiern vieler Vögel und Reptilien) stellen die Nervenzellen mit Längen bis in den Meterbereich dar; als kleinste Zellen gelten die Rickettsien und Mycoplasmen mit Durchmessern von ≈100 nm. Es ist unwahrscheinlich, daß es noch wesentlich kleinere Lebewesen geben sollte, da die untere Größe der Zelle durch die Anzahl und die Größe der Komponenten bestimmt wird, die zu einer autonomen Existenz unbedingt erforderlich sind. Dazu sind nach Schätzungen mindestens 100 Stoffwechselreaktionen notwendig. Berechnet man näherungsweise das Volumen der hierfür benötigten Enzyme, Baustoffe, Nukleinsäuremoleküle, Ribosomen, der Plasmamembran (vgl. 11.2.1.3.) und der Wassermoleküle, so ergibt sich für einen solchen „Minimalorganismus" ein Durchmesser von ≈70 nm. Um einen geregelten Ablauf der Lebensprozesse zu gewährleisten, müssen außerdem viele Molekülsorten in großer Anzahl vorhanden sein.

Eine obere Grenze für die normale Zellgröße wird durch das mit wachsendem Durchmesser immer ungünstiger werdende Verhältnis von Volumen und Oberfläche gegeben, wodurch eine hinreichend effektive Stoffaufnahme und -abgabe erschwert wird. Auch der Stofftransport sowie der Informationsfluß innerhalb der Zelle werden zunehmend problematischer.

Somit werden für Prokaryoten typische Zellgrößen von ≈1 µm beobachtet, während für Eukaryoten mit ihrer vielfältigen Kompartimentierung offenbar Größen zwischen ≈10 µm und 100 µm optimal zu sein scheinen.

11.2.1.3. Zellorganellen. Ein Überblick über die wichtigsten *Zellorganellen* einer Euzyte wird in Abb. 11.2.1–1 gegeben, wo die für Pflanzen- und Tierzellen spezifischen Bestandteile der Einfachheit halber zusammen eingezeichnet wurden. Ganz allgemein wird ein nichtplasmatisches Kompartiment vom Grundplasma durch eine *Membran* abgegrenzt.

Membranen. Je nach dem Organell, das sie umschließen, können Membranen in Aufbau und Funktion merklich differieren. Dennoch läßt der prinzipielle Bau aller Biomembranen viele Gemeinsamkeiten erkennen. Sie haben typischerweise eine Dicke von 6 bis 10 nm und bestehen vorwiegend aus polaren *Lipiden* und Proteinen (Eiweißen, vgl. 11.2.2.2.). Dabei bilden die Lipide (besondere Bedeutung kommt hier den *Phospholipiden* zu) eine Doppelschicht in der Art, daß die *polaren hydrophilen* (wasserliebenden) Kopfgruppen nach außen zum umgebenden wäßrigen Medium weisen, während die *aliphatischen hydrophoben* (wasserabweisenden) Kohlenwasserstoffketten das Innere bilden. Die Lipidmoleküle (ebenso wie verschiedene Membranproteine) sind i. allg. relativ beweglich, was zur Vorstellung von der Membran als *„flüssiges Mosaik"* (engl. fluid mosaic, Singer und Nicolson 1972) aus Lipiden und Proteinen führte.

Obwohl relativ stabile Membranmodelle *ohne* Proteine schon gewisse Funktionen der Abgrenzung und selektiven Durchlässigkeit für bestimmte Stoffe simulieren können, sind doch gerade die Proteine für sehr viele spezifische Aufgaben von entscheidender Bedeutung (z. B. aktiver Transport, Energiebereitstellung u. a.). Allen Membranen gemeinsam sind die Funktionen der *Diffusionsbarriere*, der selektiven Permeabilität (Durchlässigkeit) sowie die Fähigkeit zum spezifischen Transport (Aufrechterhaltung) des für die betreffende Funktion notwendigen *inneren Milieus*).

Die die Zelle nach außen hin abgrenzende Membran *(Plasmalemma)* umschließt die *nukleozytoplasmatische Matrix* mit *Zellkern* und *Grundzytoplasma*. Ihm kann bei Pflanzen und Bakterien noch eine *Zellwand* aufgelagert sein, die im wesentlichen aus polymeren Kohlenhydraten (Cellulose, Pektine u. a.) besteht und der Zelle eine gewisse Steifigkeit verleiht.

Membranumschlossene Gebilde (einschließlich ganzer Zellen) können miteinander verschmelzen, was besonders effektiv in (inhomogenen) elektrischen Wechselfeldern geschieht. Dabei gelingt es u. a. Erythrozyten (rote Blutzellen) zu Zellen von bis zu 0,1 mm Größe zu verschmelzen! Wichtige praktische Anwendungsmöglichkeiten dieser relativ jungen Technik zeichnen sich bereits ab.

Generell stellen Membranen ein äußerst wichtiges und sehr stark bearbeitetes Gebiet der gegenwärtigen biophysikalischen Forschung dar.

Zellkern. Er enthält die Hauptmenge der „Erbsubstanz" *Desoxyribonukleinsäure* (im folgenden nach der engl. Bezeichnung *d*eoxyribo-*n*ucleic *a*cid mit DNA abgekürzt, vgl. 11.2.2.1.) und kontrolliert damit alle Lebensprozesse. In der unstrukturierten Grundsubstanz, dem *Karyoplasma*, liegt diese als *Chromatin* vor, wo sie mit vorwiegend basischen Proteinen (Histonen, Protaminen) assoziiert ist. Erst bei der Kernteilung kondensiert das Chromatin zu den lichtmikroskopisch beobachtbaren *Chromosomen*. Auffällig sind schon im lichtmikroskopischen Bild dichtere, etwa kugelige Einschlüsse mit relativ hohem *Ribonukleinsäure-* (RNA) Gehalt, die als *Nukleoli (Kernkörperchen)* bezeichnet werden. Der Zellkern ist von einer Doppelmembran umhüllt, die jedoch Poren aufweist. Außerdem steht die äußere Kernmembran in direkter Verbindung mit dem rauhen *endoplasmatischen Retikulum*. Entfernt man den Zellkern, so stirbt die Zelle – ein überzeugender Hinweis auf dessen Rolle als Steuerzentrale.

Endoplasmatisches Retikulum (ER). Es stellt ein System von Doppelmembranen dar, die ein zusammenhängendes Netzwerk von kleinen Zisternen, Röhren und Vesikeln aufbauen, teilweise aber auch nur vorübergehend kommunizieren oder auseinander hervorgehen. Vielfach (im sogenannten *„rauhen ER"*) sind diese dicht mit *Ribosomen* (s. u.) besetzt.

Golgi-Apparat. Er läßt sich lichtmikroskopisch vor allem in solchen Zellen gut nachweisen, die der Stoffkonzentration und der Sekretbildung dienen (z. B. Drüsenzellen). In tierischen Zellen besteht er aus Stapeln von 5 bis 20 flachen, membranumgrenzten Säckchen („Golgi-Zisternen").

Mitochondrien sind die zwischen ≈1 und 5 µm langen und ca. 1 µm dicken Organellen der Zellatmung und Energiegewinnung. Sie besitzen eine Außen- und eine Innenmembran. Letztere hat eine durch zahlreiche leisten-, säckchen- oder röhrenförmige Einstülpungen in den Mitochondrieninnenraum (die sogenannte *mitochondriale Matrix*) erheblich vergrößerte Oberfläche. An der inneren Mitochondrienmembran befinden sich die Enzyme der *Atmungskette* (vgl. 11.3.2.3.), im Matrixraum die der Fettsäureoxidation sowie des Zitratzyklus.

Abb. 11.2.2–1 Chemischer Aufbau eines DNA-Polynukleotidstranges (rechts Schema)

Chloroplasten sind in den grünen Pflanzen die Orte der *Fotosynthese* (vgl. 11. 3.2.2.). Der Chloroplast ist von der Chloroplastenmembran umgeben und von flachen Membransystemen, den *Thylakoiden*, durchzogen, die als Stroma- oder Granathylakoide vorkommen. Die letzteren bilden flache Membranstapel (Grana), die lichtmikroskopisch nachweisbar sind. In den Thylakoidmembranen ist der grüne Farbstoff *Chlorophyll* (vgl. 11.3.2.2.) lokalisiert, der für die Fotosynthese von größter Wichtigkeit ist.

Zytosol (Hyaloplasma). Das ist der Bestandteil des Grundplasmas der Zelle, in dem mit den üblichen licht- und elektronenmikroskopischen Methoden keine deutlichen Strukturen mehr aufgelöst werden können (neuere Untersuchungen von PORTER u. a. legen jedoch auch für das Zytosol eine netzartige Struktur, das sogenannte „Mikrotrabekularsystem" nahe).

Ribosomen treten entweder „frei" im Zytosol auf (wahrscheinlich fixiert durch das Mikrotrabekularsystem) oder sind am ER angeheftet (rauhes ER). Sie haben einen Durchmesser von ≈ 15 bis 25 nm und bestehen aus ribosomaler RNA (rRNA) und Proteinen und unterscheiden sich hinsichtlich Zusammensetzung und Größe für Pro- und Eukaryoten. Beide Ribosomentypen setzen sich aus je einer kleinen und einer großen Untereinheit zusammen, für deren Aneinanderheften noch Mg^{2+}-Ionen benötigt werden. Häufig sind die Ribosomen zu *Polysomen*, d. h. zu Verbänden von vielen Ribosomen aggregiert.

An den Ribosomen erfolgt die Biosynthese der Proteine, weshalb sie auch oft als die „Eiweißfabriken der Zelle" apostrophiert werden.

Daneben enthält die Zelle natürlich noch eine Reihe weiterer wichtiger Bestandteile, z. B. Mikrotubuli, Lysosomen u. a., die für das Funktionieren der Zelle von großer Bedeutung sind.

11.2.2. Molekulare Basis der Lebensfunktionen

In diesem Abschnitt werden die *Nukleinsäuren* und die *Proteine* als die eigentlichen Träger der meisten Lebensfunktionen behandelt. Dabei repräsentieren die Nukleinsäuren, namentlich die Desoxyribonukleinsäure (DNA), die genetische Information, während die Proteine zahlreiche Aufgaben als Biokatalysatoren *(Enzyme)*, beim Transport von bestimmten Stoffen (z. B. Sauerstoff und Kohlendioxid durch Hämoglobin), bei der Bewegung (Muskelproteine) u. a. erfüllen.

Die Biosynthese solcher Makromoleküle erfolgt immer mit Hilfe von Enzymen, die selbst Proteine sind, aus kleineren Einheiten (Monomeren; bei Nukleinsäuren *Nukleotiden*, bei Proteinen *Aminosäuren*) durch Wasserabspaltung *(Polykondensation)* zu langkettigen Polymeren. Zwar kommen auch polymere Kohlenhydrate (z. B. Cellulose oder Chitin, bzw. Reservestoffe, wie Stärke oder Glykogen) in nicht unbeträchtlichem Maße in Lebewesen vor, doch sind diese im Gegensatz zu Nukleinsäuren und Proteinen, die aus einer definierten Bausteinzahl in genau festgelegter (und insbesondere bei Proteinen für die spezifische Funktion entscheidender) unverzweigter Anordnung aufgebaut sind, aus gleichartigen Monomeren zusammengesetzt. Die Kettenlängen sind dabei gewissen Schwankungen unterworfen und auch Verzweigungen sind möglich. Damit sind nur Nukleinsäuren und Proteine aufgrund der ungeheuren Vielzahl von verschiedenen denkbaren Anordnungen der jeweiligen Grundbausteine (4 Nukleotide für DNA bzw. RNA, 20 Aminosäuren für Proteine) befähigt, Information zu tragen. So beträgt z. B. die Anzahl der theoretisch möglichen „Proteine" (Aminosäurenkombinationen) aus 200 Aminosäuren $20^{200} \approx 10^{260}$!

11.2.2.1. Nukleinsäuren. Ihre Bausteine, die Nukleotide, bestehen aus einer basischen, stickstoffhaltigen (heterozyklischen) Ringverbindung (oft kurz als Stickstoffbase bezeichnet), einem Zuckermolekül (Desoxyribose bei DNA, Ribose bei RNA) und einem Phosphatrest.

Die basischen Ringverbindungen sind die Purinbasen *Adenin* (abgekürzt mit A) und *Guanin* (G) (in DNA und RNA) sowie die Pyrimidinbasen *Thymin* (T) (in DNA) bzw. *Urazil* (U) (anstelle von Thymin in RNA) und *Zytosin* (C).

Der chemische Bau einer Nukleinsäure (DNA) ist in Abb. 11.2.2–1 dargestellt. Man erkennt das von der gleichför-

migen Struktur Zucker – Phosphat – Zucker – Phosphat – ... gebildete „Rückgrat", während die eigentliche Information in der Abfolge der Stickstoffbasen enthalten ist.

Während die RNA der Zellen meist einsträngig vorliegt, tritt die DNA immer doppelsträngig auf. Durch die berühmten Röntgen-Strukturuntersuchungen der DNA von Franklin und Wilkins und ihre geniale Interpretation durch Watson und Crick konnte gezeigt werden, daß die DNA eine Schraubenstruktur oder *Doppelhelix* ausbildet, die aus zwei sich schrauben- oder wendeltreppenartig umeinanderwindenden (Poly-)Nukleotidsträngen besteht (Abb. 11.2.2–2). Dabei stehen sich senkrecht zur Helixachse jeweils eine Pyrimidin- und eine Purinbase gegenüber, welche über *Wasserstoffbrücken* (vgl. 5.7.2.4.) miteinander wechselwirken. Es können sich nach dem Schema in Abb. 11.2.2–3 nur Adenin (A) und Thymin (T) bzw. Guanin (G) und Zytosin (C) „paaren", da nur diese Basen räumlich zusammenpassen und untereinander Wasserstoffbrücken ausbilden können, die gleichzeitig die Doppelhelixstruktur stabilisieren. Die Basenpaarung bewirkt, daß die Basenabfolge in einem Nukleotidstrang auch die des anderen, komplementären Strangs festlegt und ist somit auch die Voraussetzung für die Weitergabe der genetischen Information bei der *identischen Verdoppelung der DNA* (*Reduplikation* oder *Replikation*) im Verlaufe der Zellteilung. Die Stabilität der Doppelhelix beruht außerdem auch noch auf Van-der-Waals-Wechselwirkungen (vgl. 5.7.2.4.) zwischen den übereinandergestapelten Basen.

Die DNA-Doppelhelix stellt kein völlig statisches Gebilde dar – sie kann sich an manchen Stellen entdrillen (öffnen) und wieder schließen (Frequenzen in der Größenordnung von MHz), wodurch ein Protonenaustausch mit dem wäßrigen Milieu der Umgebung ermöglicht wird.

Im Chromatin der Eukaryoten ist die DNA mit stark basischen Proteinkomplexen (Histonen) zu den sogenannten *„Nukleosomen"* assoziiert. Jüngste Untersuchungen sprechen für die Existenz einer von der normalerweise vorkommenden rechtsgängigen Doppelhelix (B-DNA) strukturell deutlich verschiedenen linksgängigen Form (Z-DNA) (unter gewissen Randbedingungen). Wird DNA über ≈80 °C erhitzt, so kommt es zu ihrer thermischen *Denaturierung* (Auftrennung der Doppelhelix in Einzelstränge), die bei langsamer Wiederabkühlung weitgehend wieder aufgehoben werden kann, während rasche Abkühlung bleibende Denaturierung zur Folge hat. Dabei wandelt sich die Helix in ein statistisches (Zufalls-)Knäuel um (Helix-Knäuel-Übergang, engl. helix-coil-transition), was mit physikalischen Methoden (z. B. UV-Absorption) verfolgt werden kann.

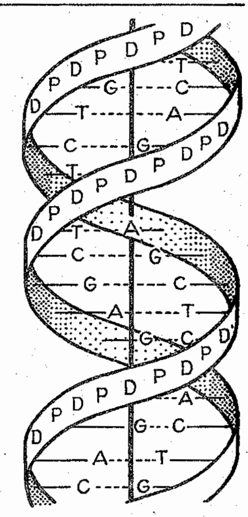


Abb. 11.2.2–2 Schematische Darstellung der Doppelhelixstruktur der DNA (P = Phosphatrest, D = Zucker [Desoxyribose])

11.2.2.2. Proteine. Die Bausteine der Proteine sind 20 natürliche α-L-Aminosäuren (L für linksdrehend; in Proteinen kommen nur linksdrehend optisch aktive Aminosäuren vor), deren allgemeine Struktur die folgende ist:

$$H_2N—\underset{\displaystyle R}{\underset{|}{CH}}—COOH,$$

wobei am sogenannten α-C-Atom sowohl die Aminogruppe —NH_2 und die Carboxylgruppe —COOH wie auch der variable Rest R hängen. R kann im einfachsten Falle ein H-Atom sein (Glyzin), aber auch längere (verzweigte und unverzweigte) Kohlenwasserstoffketten wie auch andere funktionelle Gruppen (z. B. HO—C_6H_4—CH_2— bei Tyrosin, H_2N—$(CH_2)_4$— bei Lysin) sind möglich. Im Verein mit der sauren Carboxylgruppe und der basischen Aminogruppe bestimmen die Seitenketten R die physikochemischen Eigenschaften der Aminosäuren, die somit basisch, neutral oder sauer reagieren können, hydrophil oder hydrophob sind. Dies ist natürlich auch von wesentlicher Bedeutung für ihre Rolle als Proteinbausteine.

Proteine entstehen aus Aminosäuren durch Polykondensation unter Knüpfung von sogenannten *Peptidbindungen* (—CO—NH—), wodurch ein Rückgrat des Proteins (*Polypeptids*) nach der Fom in Abb. 11.2.2–4 entsteht. Typische Größen von Proteinen liegen bei ≈100 bis 300 Aminosäuren (Molekulargewichte in der Größenordnung 10^4 bis 10^6).

Peptidbindung

Abb. 11.2.2–4 Chemische Struktur des Rückgrats eines Proteins (Polypeptids)

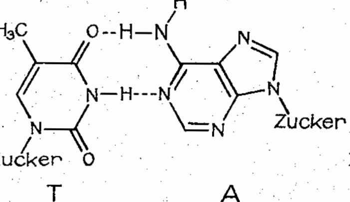


Abb. 11.2.2–3 Schema der Basenpaarung unter Ausbildung von Wasserstoffbrücken (punktiert) zwischen Thymin (T) und Adenin (A) bzw. Zytosin (C) und Guanin (G)

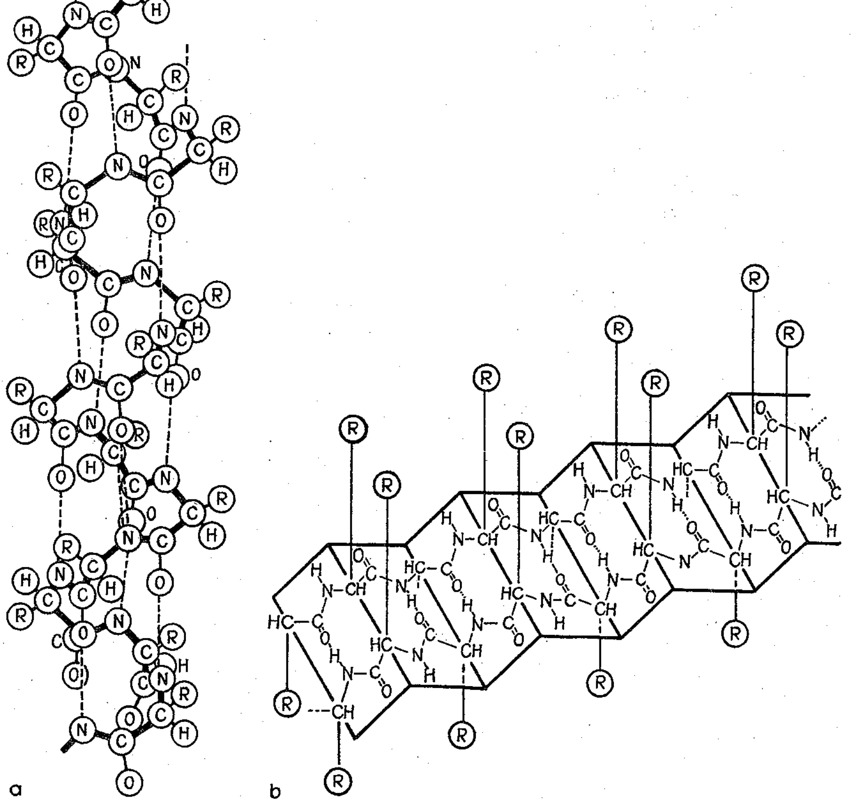


Abb. 11.2.2–5 Sekundärstrukturen von Proteinen, (a) α-Helix, (b) β-Faltblatt

Da die Peptidbindung eine *Resonanzstruktur* darstellt, liegen die Atome C_{α_1}—CONH—C_{α_2} in einer Ebene. Rotationen sind nur um die Bindungen C_α—C (Winkel Ψ) und C_α—N (Winkel Φ) relativ leicht möglich. Für bestimmte Werte von Φ und Ψ ergeben sich geordnete Überstrukturen oder *„Sekundärstrukturen"*, die man häufig als Strukturelemente in Proteinen findet. Die wichtigsten sind die α-Helix und das β-Faltblatt (Abb. 11.2.2–5).
Die für die Funktion der Proteine so wichtige *Raum-* oder *Tertiärstruktur* ist schon durch ihre Aminosäuresequenz *(Primärstruktur)* vollständig festgelegt. Dabei haben verschiedene nichtkovalente (hydrophobe, ionische, Wasserstoffbrücken u.a.) und zum Teil kovalente (Disulfidbrücken zwischen den Seitenketten der Aminosäure Zystein, die sehr reaktive SH-Gruppen enthält) Wechselwirkungen Einfluß auf die Kettenfaltung und damit die räumliche Gestalt.
Die vielleicht wichtigsten Funktionen für das Leben erfüllen die Proteine als Biokatalysatoren oder Enzyme, indem sie ganz spezifisch bestimmte Reaktionen, die sonst unter physiologischen (d.h. im Körper herrschenden) Bedingungen gar nicht ablaufen würden, durch eine drastische Herabsetzung der *Aktivierungsenergie* entscheidend beschleunigen. So erhöht z.B. das Enzym Amylase, das Stärke spaltet, die Reaktionsgeschwindigkeit bei Normalbedingungen um den Faktor 10^{11}!
Wesentliche Voraussetzung für die Enzymwirkung ist das stereochemische „Ineinanderpassen" von Enzym und Substratmolekül (das zu „verarbeitende" Molekül), wodurch ein sogenannter Enzym-Substrat-Komplex entsteht, in welchem das Substrat in einer solch günstigen Weise zu den funktionellen Gruppen des Enzyms (unter konformativer und elektronischer Destabilisierung des Substrats) angeordnet ist, daß chemische Bindungen leicht gespalten bzw. neu geknüpft werden können. Sobald das Produkt gebildet worden ist, wird es unter Wiederherstellung des Ausgangszustandes des Enzyms freigesetzt.
Daneben haben Proteine aber auch noch zahlreiche andere Funktionen, z.B. als Hormone (Insulin, MSH, Calmodulin u. a.), als Transportproteine (Carrierproteine und Ionenpumpen im Membrantransport, Sauerstofftransport durch Hämoglobin und Myoglobin), als kontraktile Proteine bei Bewegungsprozessen (z. B. Muskelaktin und -myosin), als Strukturproteine in Geweben und Hüllen (z.B. Kollagen in Sehnen, Keratin in Wolle und Haaren), bei der Abwehr körperfremder Stoffe (Immunglobuline) und als regulatorische Proteine bei der Steuerung der Aktivität anderer Makromoleküle (Enzyme, Nukleinsäuren).
Auch die Proteine sind in sich nicht völlig starr, was gerade in den letzten Jahren insbesondere auch mit Hilfe verschiedener spektroskopischer Methoden (NMR, EPR u.a.) nachgewiesen werden konnte.
Die Biosynthese der Proteine erfolgt an den Ribosomen, wo über eine Art „Negativkopie" eines der beiden DNA-

Stränge, die Boten- oder mRNA, die Nukleinsäureschrift in die Aminosäuresequenz nach den Regeln des *genetischen Codes* (nach dem eine Dreiersequenz von Nukleotiden eines mRNA-Stranges, ein sogenanntes „Basentriplett", einer Aminosäure entspricht) übersetzt wird.
Die Aufklärung der Struktur und der molekularen Wirkungsmechanismen der Proteine, vor allem der Enzyme, d. h. des Zusammenspiels von Struktur und Funktion sowie der allgemeinen Prinzipien der Regulation und Kontrolle ihrer Aktivität mit den verschiedensten physikalischen und chemischen Methoden gehört zu den wichtigsten Forschungsaufgaben der molekularen Biophysik.

11.3. Bereitstellung und Verwertung von Energie in der belebten Welt

11.3.1. Energiefluß in der Biosphäre

Grundlage allen Lebens auf der Erde ist die in der Sonne freiwerdende Energie, von der die Erde einen Teil als Strahlungsenergie erhält. Der Spektralbereich um $\lambda = 650$ nm (λ = Wellenlänge) führt nicht nur zur Erwärmung und damit zum Anstieg der Entropie, sondern wird bei der Fotosynthese in hochwertige chemische Energie umgesetzt. Dabei entsprechen 1 Mol ($= 6{,}022 \cdot 10^{23}$) Lichtquanten = 1 Einstein von 650 nm Wellenlänge einer Energie von ≈ 180 kJ.
In der Biosphäre werden zunächst aus den „energiearmen" und „entropiereichen" Verbindungen CO_2 und H_2O in den grünen Pflanzen Kohlenhydrate (Zucker und deren Polymere) und Sauerstoff erzeugt, die die Grundlage für die tierische Energiegewinnung durch Atmung (Oxidation) und damit für den Aufbau anderer „energiereicher" Biomakromoleküle, zur Verrichtung von Nutzarbeit sowie zur Wärmeproduktion bilden (Abb. 11.3.1–1).

11.3.2. Mechanismen der Bioenergetik

11.3.2.1. Allgemeine Grundlagen – Rolle des ATP. Für Lebewesen kommen als energieliefernde Reaktionen nur solche in Frage, die unter den gegebenen quasi-isothermen Bedingungen unter Abnahme der *freien Enthalpie G* (vgl. 2.6.8.5.) verlaufen. Für eine derartige „freiwillig" ablaufende *exergone Reaktion* muß die Änderung der freien Enthalpie $\Delta G = \Delta U + P \cdot \Delta V - T \cdot \Delta S = \Delta H - T \cdot \Delta S$ (ΔU = Änderung der inneren Energie, P = Druck, ΔV = Volumenänderung, T = absolute Temperatur, ΔS = Entropieänderung, ΔH = Enthalpieänderung) zwischen Ausgangs- und Endzustand stets negativ sein (vorausgesetzt, die zum Ingangbringen der Reaktion notwendige Überwindung der Aktivierungsenergie ist erfolgt, was in biologischen Systemen stets durch Enzyme wesentlich erleichtert wird). Da die meisten biologischen Prozesse bei konstantem Druck und konstanter Temperatur vor sich gehen, gibt die Änderung der freien Enthalpie (da ΔG bereits alle zusätzlichen reaktionsbedingten Umsetzungen beinhaltet) die Energie an, die maximal in Form von Arbeit aus der Reaktion entnommen werden kann.
Nun laufen in jeder Zelle natürlich auch zahlreiche *endergone*, energieaufwendige Prozesse ($\Delta G > 0$) ab. Woher stammt die Energie für all diese Vorgänge? Offenbar muß es bestimmte Speichersubstanzen chemischer Energie geben, die in fast universeller Weise zum Antreiben aller endergonen Stoffwechselvorgänge dienen können – die wich-

Adenin, Ribose, Adenosinmonophosphat (AMP), Adenosindiphosphat (ADP), Adenosintriphosphat (ATP)

Abb. 11.3.2.-1 Aufbau des Adenosintriphosphats (ATP). Die Tilde (~) kennzeichnet die sogenannten „energiereichen" Bindungen

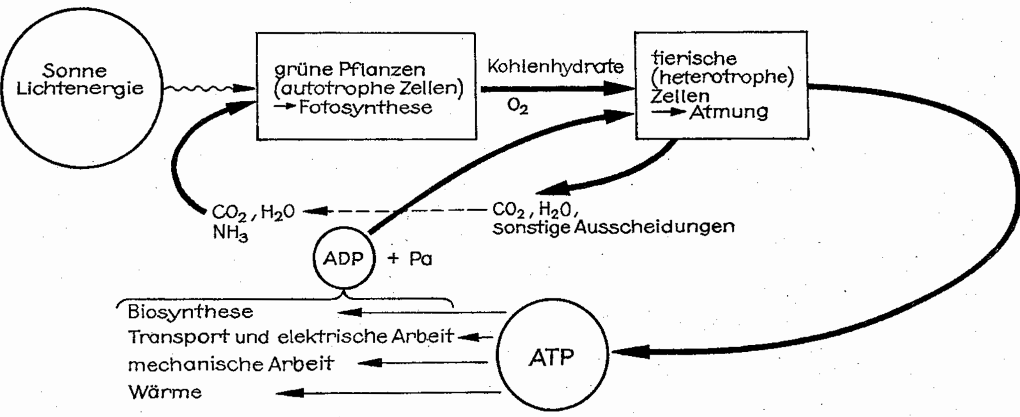


Abb. 11.3.1–1 Energiefluß in der Biosphäre

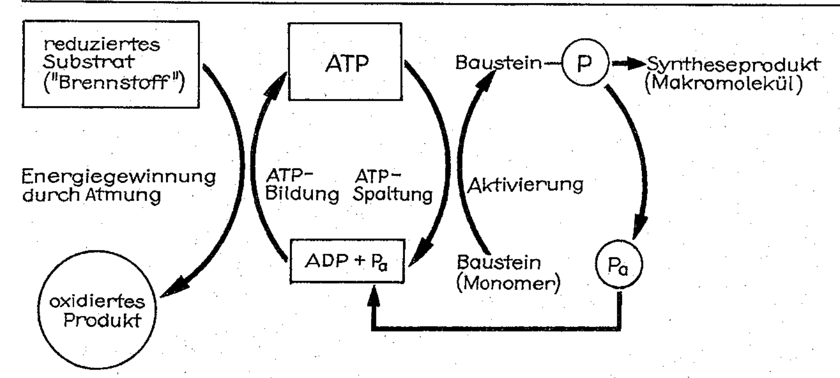


Abb. 11.3.2–2 Schema der Kopplung von energieliefernden (exergonen) und energieverbrauchenden (endergonen) Prozessen durch ATP. Durch Übertragung einer Phosphatgruppe (hier mit P oder Pa bezeichnet) wird eine Verbindung aktiviert

tigste dieser Substanzen wurde 1929 von LOHMANN mit dem *Adenosintriphosphat (ATP)* gefunden, dem Triphosphat des Nukleodids Adenosin (dessen Monophosphat gleichzeitig ein RNA-Baustein ist). Das ATP (Abb. 11.3.2–1) als universellste der Speichersubstanzen verfügt über zwei sogenannte *energiereiche Bindungen*. „Energiereich" bedeutet hier, daß bei der hydrolytischen Abspaltung der endständigen Phosphatgruppen ein relativ hoher Betrag an freier Enthalpie freigesetzt wird. Für die überwiegend vorkommende Abspaltung der letzten Phosphatgruppe kann die Reaktion schematisch folgendermaßen dargestellt werden:

$$(ATP)^{4-} + H_2O \rightarrow (ADP)^{3-} + (HPO_4)^{2-} + H^+,$$
$$\Delta G = -30\ \text{kJ/mol}.$$

Die abgespaltete Phospatgruppe wird dabei auf eine andere Verbindung übertragen, die dadurch energiereicher, d. h. aktiviert oder energetisch „aufgeladen" wird, so daß diese jetzt ihrerseits Reaktionen eingehen kann, die ohne diese Aktivierung thermodynamisch nicht möglich wären. Durch ATP können also energieliefernde und energieaufwendige Prozesse miteinander gekoppelt werden (Abb. 11.3.2–2).

11.3.2.2. Fotosynthese der grünen Pflanzen. Vermittels der Fotosynthese bauen die Pflanzen selbst unter Nutzung der Lichtenergie die energiereichen Verbindungen (vor allem Zucker, wie Glukose und deren Polymere) auf, die sie dann beim Abbau durch Atmung zur Gewinnung von chemischer Energie für die verschiedensten Lebensfunktionen wieder umsetzen. Deshalb werden sie auch als *autotrophe* Lebewesen bezeichnet.

Die Orte der Fotosynthese sind die *Thylakoidmembranen* der *Chloroplasten*. Diese enthalten viele Tausende „lichtsammelnder" und „energieleitender" Einheiten, die *Fotosysteme* oder *Antennensysteme* genannt werden, welche jeweils ≈300 *Chlorophyllmoleküle* sowie andere Hilfspigmente (z. B. 200 Karotinoide) enthalten. Die als Kollektorpigmente dienenden Chlorophyllmoleküle leiten die absorbierte Strahlungsenergie (in Form eines *Exzitons*) zu einem fotochemischen Reaktionszentrum, das durch ein an ein Protein gebundenes Chlorophyll-a_1-Molekül repräsentiert wird (sogenannter P 700-Komplex, da dessen Absorptionsmaximum bei 700 nm liegt), wodurch dieses angeregt und ein Elektron auf ein höheres Energieniveau angehoben wird. Dieses Elektron kehrt nun nicht gleich in den Grundzustand zurück, sondern es wird durch die Thylakoidmembran hindurch auf ein Elektronenakzeptormolekül (Z), dessen Natur noch nicht völlig klar ist, übertragen und von dort aus über eine Reihe weiterer Redoxkörper (Elektronenüberträger) schließlich auf den letzten Elektronenakzeptor $NADP^+$

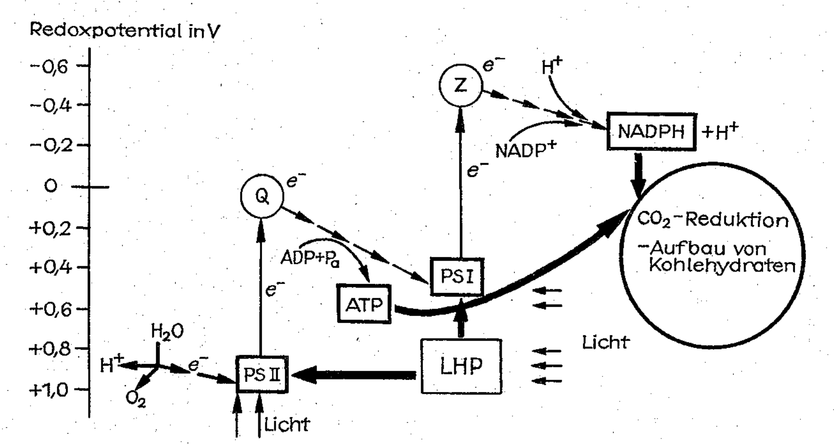


Abb. 11.3.2–3 Schema des Elektronentransportes bei der Fotosynthese. LHP = *Light harvesting protein*, Q = primärer Elektronenakzeptor des PS II, Z = primärer Elektronenakzeptor des PS I

(Nikotin-Adenin-Dinukleotidphosphat), während das oxidierte (positiv geladene) P-700-Molekül durch einen Elektronenspender wieder in seinen Ausgangszustand zurückversetzt (reduziert) wird. Das fehlende Elektron des Chlorophylls aus diesem Spender-Fotosystem I (PS I) wird aus einem vorgeschalteten Fotosystem II (PS II) ersetzt, indem von hier ein Elektron aus einem angeregten Zustand (das Absorptionsmaximum dieses Komplexes liegt bei 680 nm, deshalb auch P 680 genannt) über eine weitere Elektronentransportkette auf das PS I (P 700) übertragen wird (Abb. 11.3.2–3). Die Energiedifferenz zwischen dem angeregten Zustand des PS II und dem Grundzustand des PS I wird zur Bildung von ATP aus ADP und anorganischem Phosphat (P_a) ausgenutzt. Das dadurch im PS II hervorgerufene Elektronendefizit wird gedeckt, indem unter der Wirkung spezifischer Enzyme eine *fotolytische Wasserspaltung* erfolgt, $2\,H_2O \rightarrow 2\,H^+ + 2\,(OH^-)$, wobei die Hydroxylionen ihre Elektronen an das Chlorophyll des PS II abgeben und sich wieder zu Wasser zusammenlagern, während der überschüssige Sauerstoff in die Umgebung abgegeben wird. Eines der Wasserstoffionen (Protonen) sowie zwei Elektronen (aus der Elektronentransportkette des PS I) reagieren mit $NADP^+$ zu dem als „primäres Reduktionsäquivalent" bezeichneten NADPH (vgl. Abb. 11.3.2–3).

Die in den als *Lichtreaktion* bezeichneten oben beschriebenen Prozessen gebildeten energiereichen Produkte NADPH und ATP dienen in der *Dunkelreaktion* (die keine Lichtenergie benötigt) zum Aufbau von Kohlenhydraten aus CO_2 und H_2O in einer Reihe von enzymkatalysierten Schritten (auch als *Calvin-Zyklus* bekannt).

Zur Abschätzung des Wirkungsgrades der Fotosynthese ist davon auszugehen, daß zur Bildung von je einem Molekül NADPH während der Primärreaktionen der Fotosynthese je zwei Elektronen in den in Reihe geschalteten Fotosystemen vom Grundzustand in den angeregten (Singulett-)Zustand über eine Potentialdifferenz von jeweils ≈ 1 V angehoben werden müssen. Dazu sind jeweils zwei Lichtquanten notwendig. Zur Synthese von einem Mol Glukose (für die 12 Mol NADPH erforderlich sind) werden also mindestens 48 Einstein benötigt; da jedoch ein Teil des erzeugten NADPH und ATP unmittelbar in den Zellstoffwechsel (d.h. auch für andere Stoffwechselwege) einbezogen wird, muß man dafür ≈ 60 Einstein des energieärmeren Rotlichts von 680 nm (im günstigsten Falle) veranschlagen, d. h. 60×176 kJ = 10 560 kJ. Das ergibt, bezogen auf die 2 870 kJ, die bei der vollständigen Verbrennung eines Mols Glukose gewonnen werden, einen Wirkungsgrad von ≈ 27 %. Dieser Wert verringert sich noch beträchtlich, wenn man die Strahlungsverluste in der Atmosphäre sowie die spektrale Verteilung der Sonnenstrahlung und die Absorptionscharakteristika der Pigmentmoleküle berücksichtigt, nämlich auf knapp 8 % (ausgehend von der die Erdoberfläche erreichenden kurzwelligen Gesamtstrahlung zwischen 290 und 4 000 nm).

Der molekulare Mechanismus der ATP-Erzeugung ist in vielen Details noch unklar. Weitgehend durchgesetzt hat sich jedoch mittlerweile (gestützt durch zahlreiche experimentelle Befunde) die Auffassung, daß ein Konzentrationsgefälle von H^+-Ionen zwischen Innen- und Außenseite der Thylakoidmembran, verbunden mit einem elektrischen Feld über die Membran (durch Wasserspaltung und Elektronentransport) die wesentliche Voraussetzung für die ATP-Synthese ist (chemiosmotische Theorie, MITCHELL 1961). Neben der Erforschung der elementaren Prozesse bei der Fotosynthese werden auch Untersuchungen an künstlichen (Modell-)Systemen durchgeführt, die Teilprozesse der Fotosynthese, wie fotolytische Spaltung von Wasser in Wasserstoff und Sauerstoff oder die Ladungstrennung über eine Membran, nachzuvollziehen versuchen, da sie im Hinblick auf eine mögliche Nutzung als zusätzliche Energiequelle eine gewisse Bedeutung erlangen könnten.

11.3.2.3. Atmung. Als Atmung im allgemeinen Sinne bezeichnet man den Aufschluß der in Kohlenhydraten, Fetten und Proteinen gespeicherten chemischen Energie vermittels Oxidation und deren Überführung in für die Zelle leichter handhabbare Formen chemischer Energie – vor allem eben in Gestalt des ATP. Dabei müssen von den heterotrophen Organismen diese energiereichen organischen Verbindungen (vor allem Kohlenhydrate) als vorgefertigte Nahrungsstoffe aufgenommen werden, während diese von den autotrophen selbst synthetisiert werden.

Beim oxidativen Abbau von einem Molekül Glukose zu zwei Molekülen Milchsäure CH_3—CHOH—COOH), der sogenannten Glykolyse, werden schon zwei Moleküle ATP gebildet, die für anaerob lebende Bakterien den Energiebedarf decken müssen. In eukaryotischen Zellen wird die Glukose dagegen unter Sauerstoffzufuhr vollständig zu CO_2 und H_2O verbrannt, was wesentlich effektiver ist. Dabei wird Glukose zunächst bis zur Stufe der Brenztraubensäure (Pyruvat, CH_3—CO—COOH) abgebaut, welche in *Acetyl-Koenzym A*, eine weitere wichtige Verbindung von zentraler Bedeutung im Energiestoffwechsel, überführt wird. Über Acetyl-Koenzym A mündet der Abbauweg (wie auch der der Fette und Proteine) in den *Zitronensäurezyklus* (auch Krebs- oder Tricarbonsäurezyklus), der häufig als „zentrale Stoffwechselmühle" apostrophiert wird. In der zyklischen Reaktionsfolge des Zitronensäurezyklus erfolgt an mehreren Stellen Abspaltung von CO_2 und Wasserstoff, der auf die als Reduktionsäquivalente bezeichneten Koenzyme NAD und FAD übertragen wird. Der Wasserstoff wird nun in der sogenannten *Atmungskette* zusammen mit seinen Elektronen (weswegen auch hier oft die Bezeichnung Elektronentransportkette verwendet wird) über eine Reihe von Zwischenträgern (Redoxkörpern), die in vier Komplexen mit jeweils mehreren Membranproteinen von näherungsweise gleichen Redoxpotentialen organisiert sind, an Sauerstoff unter Bildung von Wasser weitergegeben. Somit stellen die Atmungskettenvorgänge eine Art „gebremste Knallgasreaktion" dar.

In der Gesamtbilanz sieht der Glukoseabbau wie folgt aus:

$$C_6H_{12}O_6 + 38\,P_a + 38\,ADP + 6\,O_2 \rightarrow 6\,CO_2 + 44\,H_2O + 38\,ATP.$$

In der Zelle läuft die vollständige Verbrennung der Glukose, die in einer Kalorimeterbombe eine Energie von 2 870 kJ/mol liefert, in wäßriger Lösung bei nahezu isothermen Bedingungen in vielen Einzelschritten ab. Die 38 Moleküle ATP verkörpern eine für die Zelle nutzbare chemische Energie von $\approx 1\,140$ kJ/mol, was einem Wirkungsgrad von ≈ 40 % entspricht (wobei diese einfache Ableitung jedoch in mancher Hinsicht nicht ganz exakt ist). Die restliche Energie wird in Form von Wärme an die Umgebung abgegeben.

Die von den Elektronen bei der Übergabe von einem Komplex der Atmungskette an den nächsten (Abb. 11.3.2–4) abgegebene Energie wird zur Bildung von ATP benutzt. Auch hier soll der chemiosmotischen Theorie zufolge (vgl. 11.3.2.2.) die beim Ausgleich eines H^+-Konzentrationsgefälles über die Membran (Rückfluß der Protonen) gewon-

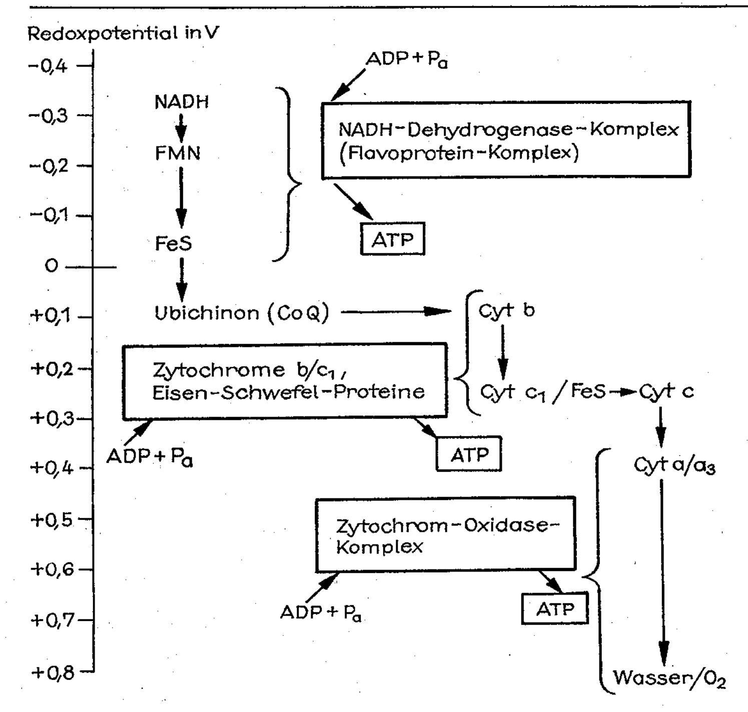


Abb. 11.3.2-4 Schematische Darstellung der Atmungskette. Die Zytochrome sind mit Cyt abgekürzt, FeS bezeichnet verschiedene Eisen-Schwefel-Proteine (der Sukzinat-Dehydrogenase-Komplex, der über Ubichinon mit dem Zytochrom b verknüpft ist, ist hier nicht dargestellt)

nene Energie zum Antreiben eines membranständigen ATP-synthetisierenden Enzyms dienen.
Letzten Endes liegt der Energiegewinnung in der Atmungskette die Bildung von Wasser aus Wasserstoff und Sauerstoff nach der Beziehung $2\,H^+ + 2\,e^- + 1/2\,O_2 \rightarrow H_2O$, $\Delta G = -235$ kJ/mol zugrunde. Diese „Knallgasreaktion" ist stark exergon und hätte, wenn sie in der Zelle tatsächlich nur in einem Schritt ablaufen würde, unweigerlich die Zerstörung derselben zur Folge. Durch die stufenweise Erniedrigung des Redoxpotentials der Elektronen in der Elektronentransportkette jedoch wird die Energie quasi „portionsweise" umgesetzt und sinnvoll zur Aufbereitung chemischer Energie in Form von ATP genutzt.

11.3.3. Energieverwertende Prozesse und ATP-Verbrauch

Der tägliche Energieumsatz des menschlichen Körpers beträgt ≈3,5 kWh. Dazu müssen ≈70 kg ATP synthetisiert und wieder gespalten werden, um die verschiedenen energieaufwendigen Vorgänge zu betreiben.
Man kann die energieverbrauchenden Prozesse grob in drei Gruppen unterteilen: chemische Arbeit zur Biosynthese, Transportarbeit und mechanische Arbeit. Elektrische Energie, die sich letzten Endes in Spannungsdifferenzen manifestiert, wird durch Unterschiede in den Ionenkonzentrationen (meist zwischen den Innen- und Außenseiten von Membranen) mit Hilfe von sogenannten Ionenpumpen, also durch Transportarbeit, erzeugt. Es gibt auch einige Tiere, die mit Hilfe von speziellen Enzymen (Luziferasen, z.B. bei Leuchtkäfern) chemische Energie (des ATP) direkt in Lichtenergie umwandeln können.
Der relative Anteil der einzelnen Energieformen am Gesamtumsatz ist von Zelle zu Zelle verschieden und hängt sowohl von den speziellen Aufgaben als auch dem Entwicklungsstand und der Aktivität der Zelle ab. So verwenden z. B. die Bakterien, deren stärkste Waffe im „Kampf ums Dasein" ihre hohe Vermehrungsrate ist, den größten Teil der Energie zum Wachstum und zur Teilung, d. h. zur Biosynthese.
Der Grundumsatz der höheren Tiere zeigt eine interessante Gewichtsabhängigkeit. Es gilt dafür näherungsweise die Beziehung $W = 3{,}37\,m^{0{,}75}$ Watt, wobei W den Grundumsatz und m die Masse des Tieres in Kilogramm bezeichnen. Daraus folgt, daß kleinere Tiere einen deutlich höheren spezifischen Grundumsatz zeigen als größere (erklärlich z. B. durch relativ größere Wärmeverluste infolge des ungünstigeren Oberfläche-Volumen-Verhältnisses).
Die Zellen höherer Tiere sind außerordentlich differenziert. So müssen sich z.B. Haut- und Darmzellen, die einem großen Verschleiß unterliegen, häufig teilen (starker Biosynthesenanteil), um zerstörte Zellen zu ersetzen. Die Nervenzellen wiederum (vgl. 11.4.1.), die der Informationsverarbeitung dienen, teilen sich gar nicht, sondern verwenden 60 bis 75 % der Stoffwechselenergie für Transportarbeit (Aufbau bzw. Aufrechterhaltung hoher Ionenkonzentrationsgradienten). Muskelzellen hingegen wenden einen beträchtlichen Anteil der Gesamtenergie zur Verrichtung mechanischer Arbeit auf.

11.3.3.1. Chemische Arbeit zur Biosynthese. In vielen Fällen müssen die Bausteine von Biomolekülen erst reaktionsfähig gemacht (aktiviert) werden. Häufig geschieht das durch die Übertragung einer Phosphatgruppe des ATP (vgl. 11.3.2.1.) oder, wie bei der Aktivierung der Aminosäuren, eines ganzen Nukleotids (AMP).
Das planvolle Knüpfen einer Peptidbindung kostet insgesamt drei Moleküle ATP. Die freie Enthalpie einer Peptid-

bindung beträgt unter zellulären Bedingungen im Mittel ≈19 kJ/mol. Damit ergibt sich für den Wirkungsgrad des Energieumsatzes 19/(3 · 30) ≈ 0,21. Die Biosynthese von polymeren Kohlenhydraten unter Bildung von Glykosidbindungen (mittlere Änderung der freien Enthalpie dabei gleichfalls $\Delta G \approx 19$ kJ/mol) erfordert dagegen nur zwei Moleküle ATP (Wirkungsgrad ≈0,32). Der Grund hierfür ist in der nichtstatistischen (eben planvollen), d.h. mit einem hohen Ordnungsgrad verbundenen Bauweise der Proteine zu sehen. Dieser hohe Ordnungsgrad ist mit einer Entropieverminderung verbunden, die mit freier Energie bezahlt werden muß. Insgesamt verbraucht die Proteinsynthese den allergrößten Teil der gesamten zur Biosynthese genutzten Energie, der nach Abschätzungen von LEHNINGER für das Bakterium Escherichia coli ≈88 % beträgt.

11.3.3.2. Transportarbeit und elektrische Arbeit. Die meisten lebenden Zellen zeigen in ihrem Inneren eine gegenüber ihrem Äußeren (umgebende Flüssigkeit) deutlich verschiedene Konzentrationen zahlreicher Stoffe, insbesondere von Ionen (z. B. K^+-Ionenkonzentration im Innern einer Säugetier-Muskelzelle 39mal höher als extrazellulär). Um diese zum Teil sehr starken Konzentrationsgradienten aufzubauen bzw. aufrechtzuerhalten, bedarf es eines aktiven Ionentransports. Der *aktive Transport* von Ionen entgegen Konzentrations- und/oder Potentialgradienten ist nur unter Energiezufuhr durch ATP-Spaltung möglich. Er wird von Proteinen („molekulare Ionenpumpen") bewerkstelligt, die die Membran-Lipiddoppelschicht durchsetzen.

Zwischen Innen- und Außenseite der Membran (z.B. einer Muskelzelle) besteht als Folge des Ionen-Ungleichgewichts i. allg. eine Potentialdifferenz der Größenordnung 100 mV. Für die Konzentrationsarbeit beim gleichzeitigen Transport von einem K^+- und einem Na^+-Ion durch die Na^+-K^+-Pumpe entgegen den Konzentrationsgefällen ergibt sich damit nach $W_{\text{Konz}} = k \cdot T \cdot \ln [X]_i/[X]_a$ ($[X]_i$, $[X]_a$ = Konzentrationen der entsprechenden Ionen innerhalb bzw. außerhalb, $k = 1{,}38 \cdot 10^{-23}$ J/K = Boltzmann-Konstante) bei Benutzung der Werte 39 (K^+) bzw. 1/12 (Na^+) für die Verhältnisse $[X]_i/[X]_a$ und $T \approx 310$ K (Körpertemperatur) für den Wirkungsgrad

$$\eta = \frac{|W_{\text{Konz}}^{K^+}| + |W_{\text{Konz}}^{Na^+}|}{\Delta G_{\text{ATP}}} \approx 0{,}53$$

(mit $\Delta G_{\text{ATP}} = 5 \cdot 10^{-20}$ J). Bei diesem gegenläufigen Transport hebt sich günstigerweise die bei einem einfachen Transport noch zu berücksichtigende elektrische Arbeit auf.

11.3.3.3. Mechanische Arbeit der Muskeln. In den *Muskeln* wird chemische Energie direkt in mechanische Arbeit umgesetzt. Ein (Skelett-)Muskel ist aus einer großen Anzahl von Muskelfasern mit Durchmessern von meist 10 bis 100 µm aufgebaut, die oft so lang wie der ganze Muskel sind. Im mikroskopischen Bild zeigen sie eine periodische Querstreifung oder Bänderung. Eine solche, von einer Zellmembran umhüllte Muskelfaser wird fast vollständig durch sogenannte *Myofibrillen* ausgefüllt. Die charakteristische Bänderung wird durch eine spezifische Anordnung der beiden sogenannten „Filamente" (lange Ketten) bildenden Proteine *Aktin* und *Myosin* verursacht.

Im Ruhezustand überlappen sich dicke *Myosinfilamente* und dünne *Aktinfilamente* nur wenig. Der molekulare Aufbau von Aktin- und Myosinmolekülen ist in Abb. 11.3.3−1 dargestellt.

Die Myosinmoleküle können *Querbrücken* zu den Aktinfilamenten ausbilden, deren Abstände fünf Aktinmolekülen

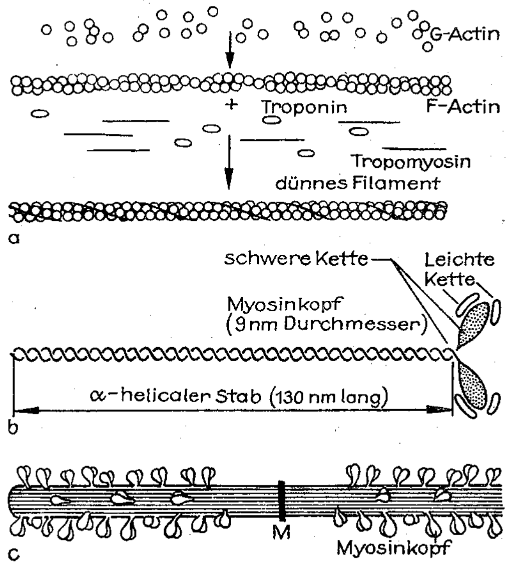


Abb. 11.3.3−1 Molekularer Aufbau (a) eines dünnen Filamentes, (b) eines Myosin-Dimeren und (c) eines dicken Filamentes

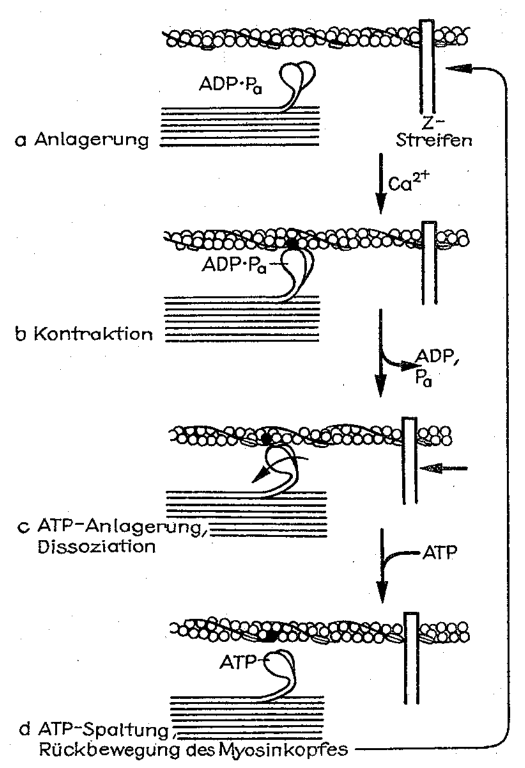


Abb. 11.3.3−2 Schematische Darstellung des Bewegungszyklus' eines Myosinkopfes während der Muskelkontraktion (Querbrückenmechanismus)

entsprechen. In Abwesenheit von Ca^{2+}-Ionen sind die Tropomyosinmoleküle, die das Troponin binden (vgl. Abb. 11.3.3–1), so über die Aktinkügelchen gespannt, daß deren Bindungsstellen für die Myosinköpfchen verdeckt werden. Wenn jedoch von den Nerven eintreffende Signale (Aktionspotentiale, vgl. 11.4.1.2.) eine Ausschüttung von Ca^{2+}-Ionen in die Zelle veranlassen, so wird im Troponin eine Konformationsänderung ausgelöst, die letzten Endes die Ursache für die Freigabe der Bindungsstellen der Aktinmoleküle für die Myosinköpfchen ist. Das durch die Bindung entstandene „Aktomyosin" kontrahiert sich durch Umklappen des Myosinkopfes um einen Winkel von $\approx 45°$ unter gleichzeitiger Freisetzung von ADP und P_a, welche aus der Hydrolyse von schon vorher gebundenem ATP stammen. Durch den Kontraktionsprozeß werden die dünnen zwischen die dicken Filamente gezogen. Nachfolgende Bindung eines weiteren ATP-Moleküls veranlaßt die Trennung von Aktin und Myosin, worauf die Querbrücke ein weiter vorn liegendes Aktinmolekül erfassen kann (Abb. 11.3.3–2).

Die Myosinköpfe (Querbrücken) führen also „ruderschlagartige" Bewegungen aus, wodurch die Aktin- in die Myosinfilamente hineingleiten (Gleitfilamenttheorie, Huxley u. a. 1953). Damit eine kontinuierliche Kontraktion (Verkürzung) des Muskels erfolgen kann, muß die Bewegung der Myosinköpfchen asynchron gleitend sein.

Bei einer Muskelkontraktion kann ein Muskel sich innerhalb von ≈ 200 ms um 20 bis 25 % verkürzen und dabei einen Druck von $\approx 4 \cdot 10^5$ Pa erzeugen, wobei der Wirkungsgrad der Umsetzung von chemischer Energie (des ATP) in mechanische Arbeit zu 40 bis 60 % abgeschätzt wurde.

11.4. Informationsaufnahme und -verarbeitung

11.4.1. Nervenzelle, Membranpotential, Erregung

Das Nervensystem bildet die Grundlage für die Leitung und Verarbeitung von Informationen. Die Bausteine des Nervensystems sind die Nervenzellen oder *Neurone*. Den Grundbauplan der Neurone, der trotz variierender Größe und Form stets gleich ist, zeigt Abb. 11.4.1–1. Das *Axon*, d. h. der Fortsatz, der die Nervenzelle mit anderen Zellen verbindet, kann Längen von wenigen Mikrometern bis zu mehr als einem Meter erreichen. Die Verbindungsstelle einer axonalen Endigung mit einer anderen Zelle wird *Synapse* genannt.

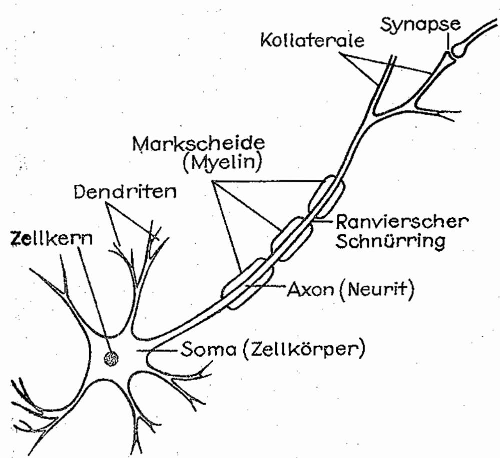


Abb. 11.4.1–1 Schematische Darstellung einer Nervenzelle (Neuron) mit ihren typischen Bestandteilen

In den Wirbeltieren ist das Axon von einer mehrschichtigen *Markscheide* aus dem Lipoprotein *Myelin* umhüllt, die in Abständen von einem bis zu wenigen Millimetern von sogenannten *„Ranvierschen Schnürringen"* unterbrochen ist.

11.4.1.1. Das Ruhepotential. Wie schon unter 11.3.3.2. kurz erwähnt, besteht zwischen dem Zellinnern und der Zellumgebung ein Ionenungleichgewicht, das letztlich auch für die Ausbildung einer Potentialdifferenz zwischen Innen- und Außenseite der Nervenzellmembran verantwortlich ist und aus historischen Gründen (eigentlich nicht ganz korrekt) als *Membranpotential* bezeichnet wird. Es hat für die Nerven- und Muskelfasern von Warmblütern i. allg. Werte zwischen ≈ -55 und -100 mV (für Säugetiere von -70 bis -90 mV), d. h. die Innenseite ist negativ gegen die Außenseite aufgeladen.

Dabei ist anzumerken, daß nur eine sehr geringfügige Nettoladungsdifferenz aller beteiligten Ionen ausreichend ist, um eine Potentialdifferenz von ≈ -80 mV über die 6 bis 10 nm dicke Axonmembran hervorzubringen, d. h. sowohl Axoninneres als auch die Zellumgebung sind im wesentlichen als elektrisch neutral anzusehen.

Ohne die Zellmembran als Diffusionsbarriere würden sich die starken Konzentrationsunterschiede zwischen Zellinnerem und -äußerem (insbesondere für K^+, Na^+, Cl^-) durch Diffusion in kürzester Zeit ausgleichen. Nun ist die Zellmembran für die genannten Ionen allerdings nicht völlig undurchdringlich. Genauere Untersuchungen zeigten, daß sie für K^+, Na^+ und Cl^- in unterschiedlichem Maße durchlässig (permeabel) ist. An dem besonders großen und damit der Messung besonders leicht zugänglichen Tintenfisch-Riesenaxon verhalten sich die relativen Permeabilitäten für K^+, Na^+ und Cl^- wie 1:0,04:0,45 (in der nichterregten Zelle), wobei spezifische Ionenkanäle in der Membran für die entsprechenden Leitfähigkeiten verantwortlich sind. Von hervorragender Bedeutung für das Ruhepotential sind dabei die K^+-Ionen, die durch die ständige Diffusion in Richtung ihres Konzentrationsgradienten (d.h. nach außen) eine Ladungsverschiebung bewirken, so daß effektiv eine positive Aufladung der Außenseite und durch den Verlust von positiven Ladungen im Innern eine negative Ladung resultieren. Die dadurch aufgebaute Potentialdifferenz wird als „Kaliumdiffusionspotential" bezeichnet. Mit dem fortschreitenden Aufbau dieses Potentials wird die (Netto-)Diffusion der K^+-Ionen natürlich durch die Anziehung der negativen Membraninnenseite zunehmend abgebremst. Wenn die durch den Konzentrationsgradienten und durch das elektrische Feld ausgeübten Kräfte betragsmäßig gleich sind, kommt die Netto-Diffusion zum Erliegen.

Berücksichtigt man nun auch noch die Diffusion von Na^+- und Cl^--Ionen entsprechend ihrer relativen Leitfähigkeiten g, so kann das Ruhepotential in guter Näherung nach einer von Goldmann aufgestellten Gleichung wie folgt berechnet werden:

$$U_M = \frac{R \cdot T}{F} \cdot \ln \frac{g_K \cdot [K^+]_a + g_{Na} \cdot [Na^+]_a + g_{Cl} \cdot [Cl^-]_i}{g_K \cdot [K^+]_i + g_{Na} \cdot [Na^+]_i + g_{Cl} \cdot [Cl^-]_a}.$$

Die passiv in die Zelle einströmenden Ionen werden durch aktiven Transport wieder aus der Zelle befördert.

11.4.1.2. Erregung – Das Aktionspotential und seine Fortleitung. Wenn das Ruhepotential einer Nerven- oder

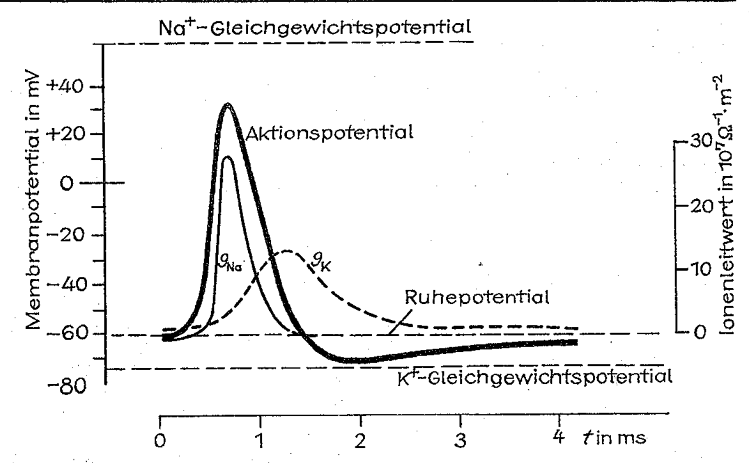


Abb. 11.4.1–2 Zeitverlauf der Ionenleitfähigkeiten während eines Aktionspotentials (nach NOBLE)

Muskelzelle durch irgendeine Einwirkung, einen *„Reiz"*, kurzzeitig um einen gewissen Betrag angehoben (d. h. positiver) und damit die Membran „depolarisiert" wird, so kommt es beim Überschreiten eines bestimmten Grenz- oder Schwellenwertes, des *„kritischen"* oder *„Schwellenpotentials"* (häufig zwischen ≈ -50 und -60 mV) innerhalb sehr kurzer Zeit ($\approx 0{,}2$ bis 0,5 ms für Nerven- und Muskelzellen von Warmblütern) zu einer Öffnung von Kanälen für Na^+-Ionen in dem Membranabschnitt, der in Richtung des sich fortbewegenden Signals *(Aktionspotentials)* unmittelbar vor dem Bereich mit der veränderten Potentialdifferenz liegt, so daß hier Na^+-Ionen in das Axon einströmen können. Damit sinkt auch in diesem Abschnitt das Membranpotential (betragsmäßig), und der nächste Bereich wird nun für Na^+-Ionen stärker durchlässig. Auf diese Weise pflanzt sich das Aktionspotential über das ganze Axon fort. Durch die mit dem Einfließen von Na^+-Ionen verbundene weitere Membrandepolarisation werden zusätzliche Kanäle geöffnet, die ihrerseits infolge des Wechselspiels von Membrandepolarisation und Na^+-Permeabilität für ein lawinenartiges Aufschaukeln dieses Prozesses sorgen, bis alle Kanäle geöffnet sind. Das Membranpotential steigt infolgedessen auf Null und erreicht schließlich sogar positive Werte (Abb. 11.4.1–2). Kurz bevor das Natrium-Gleichgewichtspotential (oder Na^+-Diffusionspotential) anliegt, bei dem effektiv kein Na^+-Einstrom erfolgt, schließen sich die Na^+-Kanäle und kurze Zeit später öffnen sich andere Kanäle, die die K^+-Ionen aus der Zelle hinausströmen lassen (vgl. Abb. 11.4.1–2). Hierdurch kehrt das Potential zu negativeren Werten und endlich zum Ruhepotential zurück. Die gegenüber dem Ruhezustand veränderte Ionenverteilung wird durch die Tätigkeit der Ionenpumpen wieder ausgeglichen.

Quantitativ können die Vorgänge bei der Ausbildung des Aktionspotentials durch die phänomenologischen *Hodgkin-Huxley-Gleichungen* beschrieben werden, welche die zeitliche Entwicklung der K^+- und Na^+-Leitfähigkeiten der Membran in Abhängigkeit vom jeweiligen Wert des Membranpotentials angeben.

Nach Ablauf eines Aktionspotentials kann die Zelle erst nach Verstreichen einer sogenannten *Refraktärzeit* ein neues Signal weiterleiten. Diese verhindert das Zurücklaufen des Impulses und das Ein- oder gar Überholen von Impulsen untereinander. Damit wird gleichzeitig die maximale Impulsfolgefrequenz auf $\approx 1\,000$/s begrenzt.

Die Übertragung der Erregung von einer Nervenzelle zu einer anderen geschieht an den Synapsen, die eine Vielzahl chemisch gesteuerter Ionenkanäle enthalten. Eine Synapse wird von einer *präsynaptischen* Struktur, z. B. einer Axonendigung, und einer *postsynaptischen* Zelle gebildet, die durch einen *synaptischen Spalt* von ≈ 20 nm Breite getrennt sind. Das am Ende der Nervenfaser eintreffende Aktionspotential bewirkt die Ausschüttung einer Überträger- oder Transmittersubstanz (sogenannte *Neurotransmitter*, z. B. Acetylcholin), die den synaptischen Spalt durchquert und nach Bindung an Rezeptormoleküle der chemisch gesteuerten Ionenkanäle in der postsynaptischen Membran Änderungen der Leitfähigkeit hervorruft, welche auch hier die Weiterleitung des Aktionspotentials ermöglichen.

Die bei Wirbeltieren vorhandene Myelinscheide um das Axon stellt einen sehr guten elektrischen Isolator dar, so daß Ionenströme nur an den Ranvierschen Schnürringen auftreten können. Somit „springen" die Nervenimpulse gleichsam von einem Ring zum nächsten *(saltatorische Erregungsleitung)*, wodurch gegenüber Nervenfasern gleicher Dicke ohne Myelinscheide wesentlich höhere Leitungsgeschwindigkeiten (bis zu ≈ 100 m/s) erreicht werden. Gleich schnelle Erregungsleitung ohne Myelinscheide würde 25 mal so starke Nervenfasern erforderlich machen.

11.4.2. Das Sehen als Beispiel für sensorische Signalverarbeitung

11.4.2.1. Das Wirbeltierauge als optisches System. Über den Gesichtssinn empfangen und verarbeiten die Wirbeltiere einen Großteil der Informationen aus und über ihre Umwelt. Das hierzu benutzte lichtempfindliche Organ ist das *Auge*, dessen schematischer Aufbau in einem Längsschnitt in Abb. 11.4.2–1 dargestellt ist.

Das optische System des Auges entwirft auf der *Netzhaut (Retina)* ein umgekehrtes und stark verkleinertes Bild der betrachteten Objekte. Den Hauptbeitrag zur Brechkraft des Auges liefert die Grenzfläche Luft – Hornhaut *(Cornea)* mit ≈ 42 Dioptrien (D). Die Erhöhung der Brechkraft des normalen Auges auf 58,6 D bei Einstellung *(Akkomodation)* auf unendlich weit entfernte Gegenstände wird durch die *Linse des Auges* vollbracht. Soll dagegen ein Objekt in einer Entfernung von ≈ 30 cm vom Auge (z. B. eine Buchseite beim Lesen) scharf gesehen werden, so muß die Brechkraft der

Augen um 1/0,3 m ≈ 3,3 D zunehmen. Dies geschieht durch eine Erhöhung der Brechkraft der Linse über eine Zunahme der Krümmung der Linsenoberfläche. Der Abstand zwischen vorderer Hornhautfläche und Netzhaut ist nämlich unveränderlich (im normalen Auge ≈24,4 mm); Abweichungen vom Normwert dieses Abstands sind die Ursachen für Weitsichtigkeit (zu geringer Abstand) und Kurzsichtigkeit (zu großer Abstand).

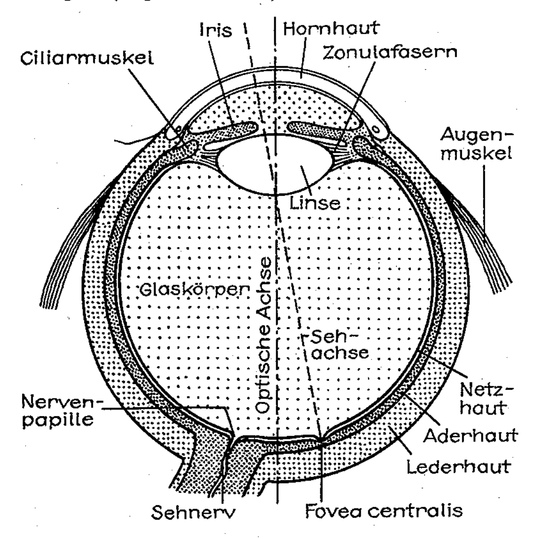


Abb. 11.4.2–1 Schnitt durch das menschliche Auge

Durch die Weite der *Pupille* wird ähnlich wie bei der Blende eines Fotoobjektivs die Menge des auf die Netzhaut gelangenden Lichts reguliert. Sie wird ihrerseits wieder reflektorisch von der Netzhaut gesteuert.

In der *Netzhautgrube* oder *Fovea centralis*, einer kleinen Eindellung mit einem Durchmesser von weniger als 1 mm, trifft die Sehachse die Netzhaut. Sie ist die Stelle des schärfsten Sehens. Dagegen kann am *„blinden Fleck"*, wo der Sehnerv die Netzhaut durchdringt, nichts gesehen werden.

Die Netzhaut stellt ein relativ kompliziertes System verschiedener Rezeptoren (Sinneszellen) und Nervenzellen dar, angeordnet in einer Rezeptor- und mehreren Nervenzellschichten. Die *Fotorezeptoren* (Lichtsinneszellen) lassen sich nach ihrer Form in *Stäbchen* und *Zapfen* unterteilen. Die Zapfen befinden sich fast ausschließlich in der Netzhautgrube bzw. in deren als *„gelber Fleck" (Macula lutea)* bezeichneten nahen Umgebung. Die über die gesamte Netzhaut verteilten Stäbchen sind wesentlich zahlreicher. Der unterschiedlichen spektralen Empfindlichkeit der drei Klassen von Zapfen (Absorptionsmaxima bei $\lambda_{max} = 440$ nm, 535 nm und 575 nm; Absorptionsmaximum der Stäbchen bei 500 nm; menschliches Auge) ist das (trivalente) Farbsehvermögen zu verdanken. Dagegen haben alle Stäbchen dieselbe spektrale Empfindlichkeit, die erheblich über der der Zapfen liegt. Deswegen sieht man bei geringer Beleuchtung (Nacht- oder skotopisches Sehen) mit den Stäbchen, bei Tageslicht hingegen mit den Zapfen (fotopisches Sehen). Da in der Fovea keine Stäbchen vorhanden sind, ist nachts ein scharfes Sehen kaum möglich. Beim Übergang vom Zapfen- zum Stäbchensehen verschiebt sich die maximale Empfindlichkeit von 560 nm nach 500 nm (Purkinje-Shift).

11.4.2.2. Aufbau eines Fotorezeptors. Das beiden Fotorezeptorarten gemeinsame Bauprinzip läßt grob eine Dreiteilung der Zelle erkennen: ein Zellabschnitt, das *Außensegment*, enthält einen Stapel sehr dichtgefalteter Membranen, ihm folgt das *Innensegment*, dann ein Zellabschnitt, der den Zellkern enthält und schließlich ein längerer Fortsatz, der mit einer Synapse als Anschluß an die nächste Nervenzelle endet.

Im Außensegment erfolgen die Lichtabsorption und die Umwandlung des Lichtreizes in ein Membranspannungssignal. Dieser Abschnitt ist vom Innensegment, das viele Zellorganellen beherbergt, räumlich getrennt.

Die Außensegmente bestehen hauptsächlich aus dichtgepackten Membranstapeln, die durch vielfache Einfaltungen der Zellmembran zustandekommen. In den Stäbchen bilden diese flache, säckchenartige Membranen, sogenannte *Disks*, die geldrollenartig übereinandergestapelt liegen. Sie werden von der Plasmamembran (Zellaußenmembran) umschlossen. Die Diskmembran enthält das lichtabsorbierende Sehpigment *Rhodopsin (Sehpurpur)*. Der Lichteinfall erfolgt in Richtung der Achse des Diskstapels (maximale Absorptionswahrscheinlichkeit).

11.4.2.3. Molekulare Prozesse bei der Verarbeitung eines Lichtreizes. Entsprechend den vier verschiedenen Fotorezeptoren existieren auch vier Sehfarbstoffe oder *Fotopigmente*, nämlich das Rhodopsin der Stäbchen und drei verschiedene Iodopsine für die drei Zapfentypen. Diese Pigmente setzen sich jeweils aus hochmolekularen Proteinen (Opsinen) und den eigentlichen farbgebenden Molekülen, den *Chromophoren*, zusammen, die bei allen vier Pigmenten identisch sind, nämlich *11-cis-Retinal*, der Aldehyd des Vitamin A (Retinol). Es ist an einen Aminosäurerest des hydrophoben, im Membraninnern befindlichen Teils des Opsins gebunden (Übergangsdipolmoment im Winkel von ≈15° zur Diskmembranebene). Das Stäbchen-Rhodopsin hat ein Absorptionsmaximum bei ≈500 nm. Wenn ein Photon durch den Chromophor absorbiert wird, so führt dies mit einer Quantenausbeute von 0,67 zur Fotoisomerisierung des gebundenen 11-cis-Retinals zu all-trans-Retinal (Abb. 11.4.2–2).

In einer Kaskade von Dunkelreaktionen wird der Erregungsmechanismus ausgelöst und schließlich die Dissoziation von Chromophor und Protein bewirkt. Dabei soll schon im ersten Schritt dieser Kaskade (zum Batho-Rhodopsin) die Isomerisierung fast bis zum all-trans-Retinal verlaufen.

11-cis-Retinal

all-trans-Retinal

Abb. 11.4.2–2 Der Chromophor Retinal in 11-cis- (oben) und all-trans-Form

Über der Plasmamembran des Stäbchens liegt im Dunkeln eine Potentialdifferenz von ≈ -30 mV (Zellinneres negativ gegen Außenraum). Ursache für das Ruhepotential sind wiederum die unterschiedlichen Permeabilitäten der Membran für K^+- und Na^+-Ionen. Dabei kommt es im Dunkeln zu einem stationären Na^+-Ionenstrom in das Außensegment, der durch einen K^+-Ionenausstrom im Innensegment kompensiert wird. Das „Zurückpumpen" der Ionen gegen ihre Konzentrationsgefälle geschieht wieder vermittels Ionenpumpen unter ATP-Verbrauch.
Nach Belichtung eines Stäbchens wird die Leitfähigkeit der Außenmembran für Na^+-Ionen verringert, wodurch der depolarisierende Einstrom von Na^+-Ionen reduziert wird und sich so das Zellinnere stärker negativ auflädt als bei Dunkelheit. Das solcherart hervorgerufene *Rezeptorpotential* besteht also in einer Hyperpolarisation der Sehzelle, deren Stärke von der Lichtintensität abhängt.
Da zur Aufrechterhaltung des stationär fließenden Ionenstromes Energie benötigt wird, wird also bei Belichtung erstaunlicherweise weniger Energie verbraucht als im Dunkeln. Abschätzungen haben ergeben, daß die Verringerung des Energiebedarfes je absorbiertes Photon mehrere Zehnerpotenzen größer ist als die Photonenenergie.
Der Mechanismus der Informationsübertragung vom Ort der Lichtabsorption (z. B. in der Mitte des Disks) zur Aussenmembran ist noch nicht völlig geklärt. Nach neueren Untersuchungen dürften sowohl Ca^{2+}-Ionen als auch zyklisches Guanosin-Monophosphat als „Transmitter" eine wesentliche Rolle spielen.
Es konnte eindeutig nachgewiesen werden, daß ein einziges je Stäbchen absorbiertes Photon zu signifikanten Abnahmen des Dunkelstromes (um mehr als 3 %) führt (Ratten- und Krötenretina). Die Retina des menschlichen Auges ist in kleine *rezeptive Felder* strukturiert. Werden gleichzeitig (d. h. innerhalb von ≈ 20 ms) 5 bis 8 Photonen der Wellenlänge 500 nm in einem rezeptiven Feld absorbiert, so wird dieser Lichtreiz bei völliger Dunkeladaptation (-anpassung) schon wahrgenommen. Folglich stellt das Auge eine Art „Koinzidenzschaltung" dar.

11.5. Physikalische Grundlagen der biologischen Selbstorganisation

Unter Selbstorganisation versteht man nach Ebeling einen irreversiblen Prozeß, der durch das kooperative Wirken von Teilsystemen zu komplexeren Strukturen des Gesamtsystems führt. Diese Form der Selbstorganisation wird häufig auch als *dissipative Selbstorganisation* (Bildung *dissipativer Strukturen*) bezeichnet, im Unterschied zur *konservativen Selbstorganisation* (z. B. bei der Kristallbindung). *Evolution* kann dann als eine unbegrenzte Folge von Prozessen der Selbstorganisation in Richtung auf eine optimale Anpassung an die Umwelt aufgefaßt werden.
In *offenen Systemen*, wie sie ja biologische Objekte darstellen, denen von außen ständig Energie und Stoff zugeführt werden, können sich stationäre *Nichtgleichgewichtszustände* mit hohem Ordnungsgrad ausbilden. Wegen deren starken Abweichungen vom Gleichgewicht muß das Verhalten solcher Systeme durch nichtlineare Gleichungen beschrieben werden, wobei für diese i. allg. mehrere stationäre Lösungen existieren. Über die Analyse ihrer Stabilität findet man die physikalisch relevante Lösung (stabil gegenüber Fluktuationen), während instabile Zustände ein „Aufschaukeln" der Fluktuationen (Verstärkung oder Amplifikation) zeigen, da das System bestrebt ist, in stabile Zustände überzugehen. Dabei können die Endzustände durchaus einen höheren Ordnungsgrad bzw. eine geringere Symmetrie haben. Der Übergang zu stabilen Zuständen wird als *physikalische Evolution* bezeichnet.
Zur Bewertung der Möglichkeiten der physikalischen Evolution kann die *innere Entropieproduktion* $P = \mathrm{d}S_i/\mathrm{d}t$ (S_i = Entropie im Innern des Systems) dienen. Die Schwankungen der inneren Entropieproduktion δP kann man als Summe der Schwankungen der verallgemeinerten Flüsse J_r bei konstanten verallgemeinerten Kräften $X_r(\delta_J P)$ und der Schwankungen der verallgemeinerten Kräfte bei konstanten Flüssen $(\delta_X P)$ ausdrücken, $\delta P = \delta_J P + \delta_X P$.
Physikalische Evolution ist nach dem *Evolutionsprinzip* von Glansdorff und Prigogine nur dann möglich, wenn gilt $\delta_X P \leqq 0$.
Im stationären *Fließgleichgewicht* ist $\delta_X P = 0$.
Bei der Entstehung von Ordnung kann den Oszillationen gewisser Systemparameter eine fundamentale Bedeutung zukommen, ja sie scheinen eine allgemeine Eigenschaft dissipativer Strukturen zu sein.
Auch für die molekulare Selbstorganisation von Biomakromolekülen und, damit zusammenhängend, die Entstehung des Lebens sind verschiedene physikalisch-mathematische Modelle vorgeschlagen worden. Eines der überzeugendsten stammt von M. Eigen, worin eine „Darwinsche Selektion" zwischen konkurrierenden, zu Selbstreplikation und Mutation fähigen Molekülen (RNA) postuliert wird, die auf einer späteren Entwicklungsstufe mit Proteinen kooperieren und so komplexere Strukturen bilden können („Hyperzyklen").
Darüber hinaus werden natürlich Selbstorganisationsprozesse auch auf ganz anderen Ebenen beobachtet, so z. B. in sogenannten rekonstituierten Enzymsystemen (wo unter gewissen Bedingungen Oszillationen in Substrat- bzw. Zwischenproduktkonzentrationen und in den Enzymaktivitäten festgestellt wurden) u. a., die ebenfalls mathematisch behandelt werden können.

12. Astro- und Weltraumphysik

Die *Astrophysik* als ein Teilgebiet der Astronomie beschäftigt sich mit der physikalischen und chemischen Beschaffenheit der kosmischen Objekte. Zu diesen gehören neben den Sternen auch alle nichtstellaren außerirdischen Körper, z.B. die des Planetensystems, sowie die diffus im Weltraum verteilte Materie und die dort auftretenden Felder. Der physikalische Zustand der Erde einschließlich ihrer Atmosphäre wird von den Geowissenschaften erforscht, deren Untersuchungsmethoden sich wesentlich von denen der Astronomie unterscheiden.
Im Gegensatz zu den experimentierenden Naturwissenschaften, wie Physik, Chemie, Biologie, kann die Astronomie mit ihren Forschungsobjekten nicht experimentieren, diese also nicht aktiv beeinflussen; die Astronomie kann nur beobachten. Eine begrenzte Änderung ergibt sich durch die Raumfahrt, die u. a. seismologische, mineralogische, meteorologische und chemische Untersuchungen erdnaher Himmelskörper sowie Experimente im interplanetaren Raum ermöglicht. Ansonsten muß sich die Astronomie, speziell auch die Astrophysik, auf die Untersuchung der einfallenden elektromagnetischen und Korpuskularstrahlung hinsichtlich Richtung, Intensität, spektrale Zusammensetzung sowie Polarisation beschränken. Diese Beobachtungen werden theoretisch interpretiert und verallgemeinert, wozu die Einbeziehung von Forschungsergebnissen anderer Naturwissenschaften, vor allem von Physik, Chemie und den Geowissenschaften, notwendig ist. Man gelangt dadurch zu wissenschaftlich begründeten Aussagen über die die Strahlung aussendenden aber auch über die die Strahlung auf ihrem Weg zum Beobachter beeinflussenden kosmischen Objekte und Felder. Die theoretischen Interpretationen der astronomischen Beobachtungen führen dabei auf eine Vielzahl neuer Erkenntnisse über im Weltraum existierende Materiezustände, die zum Teil hinsichtlich ihrer Massen- und Energiedichten so extrem sind, daß sie in irdischen Laboratorien nie realisiert werden können.

12.1. Physik der Körper des Planetensystems

Zum *Planetensystem* wird die Gesamtheit der Körper gezählt, die sich um die Sonne bewegen, sowie der Raum um die Sonne, in dem die Bahnen dieser Körper liegen. Sonne und Planetensystem zusammen bilden das *Sonnensystem*.
Die Unterscheidung zwischen Sonne einerseits und dem Restsystem andererseits hat physikalische Gründe. Die Sonne ist so massereich, daß in ihrem Innern Kernprozesse ablaufen, bei denen große Energiemengen freigesetzt und schließlich an der Oberfläche abgestrahlt werden. Sie gehört damit zu den Sternen (vgl. 12.2.). Alle anderen Körper des Systems sind nicht selbstleuchtend; sie reflektieren nur das empfangene Sonnenlicht oder absorbieren es, und strahlen es danach in anderen Wellenlängenbereichen wieder aus.
Planeten und *Planetoiden* sind im wesentlichen feste Körper, die sich in Ellipsenbahnen bewegen, in deren einem Brennpunkt die Sonne steht. Die *Planetensatelliten* sind ebenfalls große feste Körper, die aber gravitativ an einen Planeten gebunden sind, also diesen umlaufen. *Kometen* bewegen sich zwar auch um die Sonne, haben aber eine wesentlich andere Struktur als die Planeten und Planetoiden. Die von den Kometen in den interstellaren Raum gelangenden gasförmigen und festen Partikeln bilden zusammen mit dem von der Sonne abströmenden Gas die *interplanetare Materie*.

12.1.1. Planeten, Planetoiden, Planetensatelliten

12.1.1.1. Theorie des Planeteninnern. Den physikalischen Zustand und die chemische Zusammensetzung der Materie im Innern eines Planeten kann man nicht direkt untersuchen. Selbst für die Erde ist dies nicht möglich. Die bisher tiefsten Bohrungen dringen nur in die obersten Erdschichten bis ≈0,16 % des Erdradius ein. Man geht deshalb i. allg. so vor, daß man sich theoretische Modelle konstruiert, bei denen bestimmte Parameter frei wählbar sind, denen immer je eine der Beobachtung zugängliche Beobachtungsgröße, eine Observable, zugeordnet werden kann. Bestimmt man mittels Beobachtungen für einen Planeten den Zahlenwert jeder Observablen, so läßt sich ein Modell dieses Planeten numerisch berechnen. Da die Zahl der unabhängigen Beobachtungsgrößen begrenzt und die Beobachtungen selbst mit mehr oder minder großen Fehlern behaftet sind, sind auch die abgeleiteten *Planetenmodelle* in vieler Hinsicht ungenau und nicht immer eindeutig. Damit sie die Wirklichkeit möglichst gut beschreiben, müssen sie ein weiteres, grundsätzliches Kriterium erfüllen: Die Modelle der Planeten müssen in ihrer Gesamtheit kosmogonisch sinnvoll sein. Dies bedeutet, daß der von ihnen beschriebene gegenwärtige Zustand das Ergebnis eines für die Körper des Sonnensystems gleich langen, aber nicht notwendig gleich verlaufenden Entwicklungsprozesses darstellen muß, wobei die Entstehung der Planeten aus einem gemeinsamen Ursonnennebel angenommen wird.
Das Modell eines Planeten stellt im Idealfall die Gesamtheit der Kenntnisse über dessen inneren Aufbau dar. Im einzelnen muß es Aussagen über die Dichte-, Druck- und Temperaturschichtung sowie über die chemische Zusammensetzung in Abhängigkeit vom Mittelpunktabstand r liefern. An astronomischen Beobachtungsgrößen stehen dazu u. a. die Planetenmasse M sowie der Äquatorradius R_A und der Polarradius R_P zur Verfügung. Der Unterschied der beiden Radien ist eine Folge der Planetenrotation, die eine Abplattung $e = (R_A - R_P)/R_A$ bewirkt. Die äußere Form des Planetenkörpers wird durch die Druckschichtung im Innern und durch die Winkelgeschwindigkeit ω (bzw. die Rotationsgeschwindigkeit) des Planeten bedingt. In erster Näherung läßt sich der Planetenkörper durch ein Rotationsellipsoid mit R_A und R_P als Äquator- bzw. Polhalbachse beschreiben. Die mittlere Dichte des Planeten ergibt sich dann zu $\bar{\varrho} = 3M/(4\pi R_P R_A^2)$. Auch sie ist eine den inneren Aufbau des Planeten charakterisierende Beobachtungsgröße.

Die näherungsweise rotationsellipsoidische Form zeigt, daß sich das Planeteninnere weitgehend wie eine rotierende Flüssigkeit verhält, die nur ihrem eigenen Schwerefeld ausgesetzt ist. Man kann daher überall im Innern *hydrostatisches Gleichgewicht* voraussetzen, was bedeutet, daß die Summe der in einem Punkt wirkenden Druck-, Gravitations- und Zentrifugalkraft gleich Null ist. Trotz der näherungsweise ellipsoidischen Form des Planetenkörpers kann man wegen der Kleinheit von ω zunächst der Einfachheit halber eine kugelsymmetrische Dichteverteilung im Planeteninnern annehmen. Es gilt dann $\varrho = \varrho(r)$ und für die Schwerebeschleunigung $g = g(r)$, wobei diese durch diejenige Masse M_r bestimmt ist, die sich innerhalb der Kugel mit dem Radius r befindet, $g(r) = GM_r/r^2$. Die Rotation bewirkt eine Zentrifugalbeschleunigung, die proportional ω^2 und dem Abstand des betrachteten Punktes von der Rotationsachse ist. Dieser wiederum ist sowohl von r als auch vom Winkel zwischen der Rotationsachse und der Verbindungslinie Mittelpunkt – betrachteter Punkt abhängig. Um allein eine r-Abhängigkeit zu haben, betrachtet man den Mittelwert der Zentrifugalbeschleunigung für die Oberfläche der Kugel mit dem Radius r. In erster Näherung ergibt sich $g_z(r) = 2/3\,\omega^2 r$. Die Gleichung des hydrostatischen Gleichgewichts kann man damit in der Form

$$\frac{dp(r)}{dr} = -([GM_r/r^2]\,\varrho(r) - 2/3\,\omega^2 r\varrho(r)) \qquad (12.1.)$$

schreiben, wobei $p(r)$ den im Abstand r herrschenden Druck bezeichnet. Sie besagt, daß die Druckänderung $dp(r)$ bei einer kleinen Radiusänderung dr gleich der im Abstand r auf die Flächeneinheit wirkenden und zum Zentrum hin gerichteten Schwerkraft ist, reduziert um die in Gegenrichtung wirkenden Zentrifugalkraft. Das Minuszeichen steht, weil mit wachsendem r der Druck sinkt.

Das Schwerefeld außerhalb eines Planeten ist durch dessen innere Dichteverteilung eindeutig bestimmt. Dies gilt aber nicht umgekehrt: Aus den beobachteten Wirkungen des Schwerefeldes auf einen anderen gravitierenden Körper, z.B. einen natürlichen oder künstlichen Satelliten, läßt sich die Dichteverteilung $\varrho(r)$ nicht eindeutig erschließen. Eine gewisse Auskunft über die innere Dichteverteilung gibt aber das aus der Bahnanalyse von Satelliten ableitbare dimensionslose Verhältnis $\alpha = C/(MR_A^2)$, in dem C das Hauptträgheitsmoment des Planeten bezeichnet. Für einen hypothetischen Planeten konstanter Dichte ist $\alpha = 0{,}4$; je stärker die Masse zum Zentrum zu konzentriert ist, um so kleiner ist α. Der Trägheitsfaktor α stellt damit eine weitere für die Theorie wichtige Beobachtungsgröße dar, nur zeigt es sich, daß ganz unterschiedliche Dichteverteilungen gleiche Werte für α liefern können. Tab. 12.1.1–1 enthält die Zahlenwerte der genannten Beobachtungsparameter für die einzelnen Planeten.

Tab. 12.1.1–1 Für den inneren Aufbau der Planeten wesentliche Beobachtungsgrößen

Planet	Äquatorradius in km	e	$\overline{\varrho}$ in kg/m³	α
Merkur	2439	0	5430	–
Venus	6052	0	5240	–
Erde	6378	0,00335	5515	0,331
Mars	3397	0,00585	3930	0,38
Jupiter	71398	0,0629	1330	< 0,26
Saturn	60000	0,109	700	< 0,26
Uranus	25400	(0,020)	1270	(> 0,3)
Neptun	24300	(0,023)	1710	(> 0,3)
Pluto	(1750)		(700)	

Da die Dichteverteilung des Planeteninnern nicht eindeutig aus rein astronomischen Beobachtungsgrößen erschließbar ist, muß man auf Erkenntnisse zurückgreifen, die anderweitig gewonnen werden. Dies hat aber zur Folge, daß die Planetenmodelle i. allg. noch große Unsicherheiten enthalten. Eine wesentliche Rolle spielt bei diesem Vorgehen die *Zustandsgleichung* der im Planeteninnern vermuteten Stoffe. Sie beschreibt den Zusammenhang von Druck, Dichte, Temperatur und chemischer Zusammensetzung und wird bei der Integration der Gleichung (12.1.) benötigt. Für kondensierte Materie ist die Zustandsgleichung sehr kompliziert und stark materialabhängig. Zu ihrer Ermittlung sind aufwendige Experimente der Hochdruckphysik notwendig. Solche Untersuchungen lassen Diskontinuitäten im Stoffverhalten im Planeteninnern erwarten, die durch Phasenübergänge verursacht werden, z. B. beim Übergang vom festen zum flüssigen Zustand, oder beim Übergang zwischen verschiedenen Hochdruckmodifikationen von Mineralen, oder beim Übergang vom molekularen zum metallischen Wasserstoff. Man hat daher mit einem Schalenaufbau der Planeten zu rechnen. Die Schalenstruktur läßt sich beobachtungsmäßig bei denjenigen Planeten und Satelliten nachweisen, bei denen seismologische Untersuchungen möglich sind, d. h. gegenwärtig bei der Erde und beim Mond. Markante Diskontinuitäten existieren auch in Form von chemischen Inhomogenitäten. Sie ergeben sich im schmelzflüssigen Zustand des Materials durch eine Trennung der verschiedenen Stoffkomponenten entsprechend ihrer Dichte unter der Wirkung des Gravitationsfeldes. Eine solche inhomogene Struktur erschwert die Berechnung von Planetenmodellen, da für jede Schicht eine andere Zustandsgleichung gilt. Man geht daher schrittweise vor, indem man die chemische Zusammensetung und die Dicke der einzelnen Schichten solange ändert, bis eine befriedigende Übereinstimmung zwischen den beobachteten und theoretisch berechneten Modellparametern erreicht ist.

12.1.1.2. Planetenatmosphären, planetare Magnetfelder. Die einen Planeten umgebende Gashülle bezeichnet man als seine Atmosphäre. Sie stellt in ihrem jetzigen Zustand das Ergebnis einiger seit der Planetenentstehung wirkender, zum Teil gegenläufiger Prozesse dar. Jeder Planetenatmosphäre geht Gas verloren, da diejenigen Atome und Moleküle, die eine höhere als die für jeden Planeten charakteristische Entweichgeschwindigkeit haben, infolge ihrer großen mittleren freien Weglänge in den äußeren Atmosphärenschichten in den Weltraum entweichen können. Die Verluste sind umso geringer, je größer die Schwerebeschleunigung und die Molekülmasse sind und je niedriger die Atmosphärentemperatur ist. Leichte Gase, wie Wasserstoff, entweichen daher bevorzugt. Durch chemische Prozesse vor allem an den Planetenoberflächen, atmosphärische Kondensationsvorgänge und gegebenenfalls biologische Aktivitäten wird Gas aus der Atmosphäre verbraucht und die chemische Zusammensetzung verändert. Bei der Entgasung des Planeteninnern, hauptsächlich in Verbindung mit Vulkanismus, werden andererseits Gase charakteristischer Zusammensetzung freigesetzt.

Die chemische Zusammensetzung sowie die Temperatur- und Dichteschichtung der Atmosphäre lassen sich direkt mit meteorologischen Mitteln, zum Teil in Verbindung mit

Raumfahrtexperimenten, ermitteln. Spektralanalytische Untersuchungen liefern i. allg. nur die chemische Zusammensetzung der äußeren Atmosphärenschichten, weil die tieferliegenden vielfach durch dichte Wolkenschichten verborgen sind. Aufschlüsse ergeben auch Experimente mit Radiostrahlung, die von Raumsonden, die sich von der Erde aus gesehen hinter den Planeten befinden, ausgesandt werden, da diese Strahlung in unterschiedlichen Atmosphärenschichten verschieden stark geschwächt wird. In den Planetenatmosphären gelten die gleichen Gasgesetze, wie sie von der Erde her bekannt sind.

Bei einigen Planeten (Erde, Merkur, Mars, Jupiter, Saturn) konnten Magnetfelder nachgewiesen werden, die in erster Näherung Dipolcharakter haben. Sie entstehen wahrscheinlich infolge eines *Dynamoeffektes*, der sich bei Konvektion in einem (mindestens teilweise) flüssigen oder gasförmigen rotierenden Himmelskörper einstellt, falls die Materie elektrisch leitend ist.

12.1.1.3. Erdartige Planeten. Die vier sonnennächsten Planeten Merkur, Venus, Erde, Mars werden aufgrund ihrer chemischen und physikalischen Eigenschaften zur Gruppe der erdartigen Planeten zusammengefaßt. Ihre mittlere Dichte (vgl. Tab. 12.1.1–1) ist deutlich höher als die der jupiterartigen Planeten Jupiter, Saturn, Uranus und Neptun. Über Pluto sind noch keine zuverlässigen Aussagen möglich.

Bei den erdartigen Planeten spielt die Erkundung ihrer Oberfläche eine besondere Rolle, da deren Beschaffenheit wichtige, weil durch Beobachtungen gewonnene Randbedingungen für die Planetenmodelle liefert. Diese Untersuchungen erfolgen nach geologischen Prinzipien. Dabei zeigt sich, daß man die Planeten entsprechend ihrer Oberflächeneigenschaften in eine Reihe mit sinkender Masse ordnen kann, wobei der Mond – obwohl er nicht zu den Planeten zählt – sich als ein Endglied in diese Sequenz einfügt: Erde, Venus, Mars, Merkur, Mond. Während die Erdoberfläche durch globale tektonische und vulkanische Aktivitäten sowie durch starke Erosionsprozesse geprägt wird, ist das Charakteristikum der Mondoberfläche jegliches gegenwärtiges Fehlen solcher Erscheinungen. Sie wird nur durch Einschläge interplanetarer Körper und von Partikeln der solaren Korpuskularstrahlung verändert. Venus liegt im Bereich zwischen den gegenwärtig magmatisch und tektonisch völlig inaktiven Körpern und der Erde.

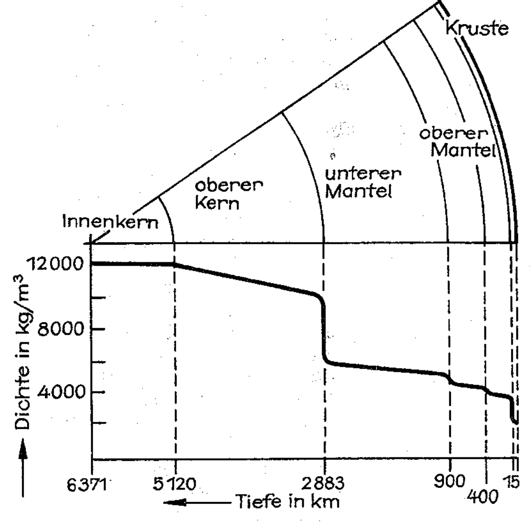
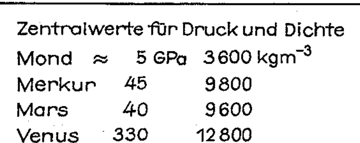
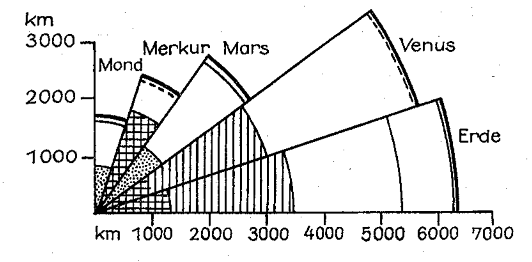


Abb. 12.1.1–2 Der Schalenaufbau der Erde und der vermutete Verlauf der Dichte im Erdkörper



Zentralwerte für Druck und Dichte

Mond ≈	5 GPa	3600 kgm⁻³
Merkur	45	9800
Mars	40	9600
Venus	330	12800
Erde	370	13000



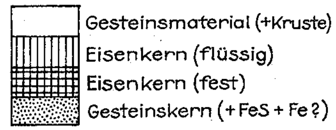


Abb. 12.1.1–3 Der Schalenaufbau der erdartigen Planeten und des Mondes

Der Schalenaufbau der Erde ist durch seismologische Untersuchungen nachgewiesen. Die *Erdkruste* ist die äußerste Schicht. Sie wird nach unten durch die Mohorovičić-Diskontinuität begrenzt, die unter den Kontinenten in einer Tiefe von im Mittel 35 km, unter den Ozeanen bei ≈5 bis 10 km liegt (Abb. 12.1.1–2). An der Untergrenze der Kruste nimmt die Dichte sprunghaft zu; es beginnt der *Erdmantel*, der in seinem oberen Teil von basischen Silikaten gebildet wird. Phasenübergänge der verschiedenen Minerale definieren weitere Unterschalen. Die Manteluntergrenze liegt in einer Tiefe von ≈2 900 km. In der sie bestimmenden Gutenberg-Wiechert-Diskontinuität springt die Dichte von ≈5 000 kg/m³ auf ≈10 000 kg/m³, gleichzeitig erhöht sich die elektrische Leitfähigkeit des Materials um mehrere Zehnerpotenzen. Es beginnt hier der *Erdkern*. Während sich im Erdmantel Erdbebenwellen in guter Näherung wie in einem festen Körper fortpflanzen, verhalten sie sich im oberen Erdkern wie in einer zähen Flüssigkeit, die wesentlich aus Eisen, Nickel, Cobalt und Schwefel bestehen dürfte. In dieser Schicht wird wahrscheinlich mittels eines Dynamoeffektes das Erdmagnetfeld aufgebaut. Unterhalb von ≈5 120 km ist der Erdkern fest. Die Temperaturverteilung im Erdinnern ist nur näherungsweise bekannt; im Zentrum dürften zwischen 4000 K und 7000 K herrschen. Die Energie stammt dabei aus radioaktiven Zerfällen langlebiger Isotope von Uranium, Thorium und Kalium. Die Stoffzusammensetzung ist, wie oben erwähnt, nicht eindeutig bestimmbar. Man berücksichtigt bei der Modellierung deshalb auch die anderweitig beobachtete Verteilung der häufigsten Elemente im Weltall (vgl. 12.2.2.4.), sowie deren Kondensierbarkeit unter den (angenommenen) Bedingungen im Ursonnennebel bei der Planetenentstehung.

Da man von den anderen erdartigen Planeten keine seismologischen Daten hat, sind die Modelle entsprechend hypothetisch. Im allgemeinen geht man aber von einem ähnlichen Schalenaufbau wie bei der Erde aus. Abb. 12.1.1–3 zeigt den möglichen Aufbau im Vergleich zur Erde. Das Verhältnis von Krusten-, zu Mantel- und Kerndicke sowie die angenommene Stoffzusammensetzung dieser Schichten ist durch die beobachtete mittlere Dichte der Himmelskörper bestimmt. Der Venusaufbau dürfte ähnlich dem der Erde sein, doch ist z. B. noch völlig offen, ob Venus auch einen festen metallischen inneren Kern hat.

Von den erdartigen Planeten hat Merkur praktisch keine Atmosphäre. Der aus Tab. 12.1.1–4 hervorgehende Unterschied in der Zusammensetzung der Venus- und Marsatmosphäre einerseits und der Erdatmosphäre andererseits ist auf die in ihr während der Erdgeschichte wirkenden, auf dem Vorhandensein von großen Mengen flüssigen Wassers beruhenden, chemischen und biologischen Prozesse zurückzuführen. Bei der Entstehung der Erde dürfte ihre Atmosphäre etwa die gleiche Zusammensetzung gehabt haben wie sie die Venus- und Marsatmosphäre noch gegenwärtig aufweisen. Abb. 12.1.1–5 zeigt den vertikalen Temperaturverlauf in der Venus- und der Marsatmosphäre.

12.1.1.4. Jupiterartige Planeten. *Jupiter, Saturn.* Die geringen mittleren Dichten (vgl. Tab. 12.1.1–1) zeigen, daß die beiden Planeten im wesentlichen aus den leichtesten Elementen, also Wasserstoff und Helium, bestehen müssen, wobei man i. allg. ein Verhältnis von 10 Wasserstoffatomen

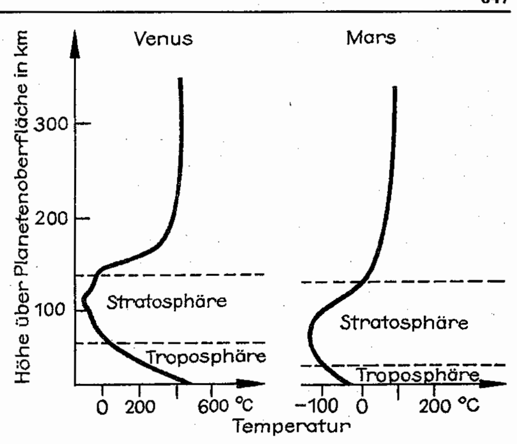


Abb. 12.1.1–5 Vertikaler Temperaturverlauf in den Atmosphären von Venus und Mars

zu 1 Heliumatom (mittlere kosmische Häufigkeit; vgl. 12.2.2.4.) annimmt. Der niedrige Wert des Trägheitsfaktors α weist aber auf eine hohe zentrale Massenkonzentration (hohe Zentraldichte) hin. Wie die Beobachtungen außerdem ergeben, strahlen beide Planeten etwa doppelt so viel Energie ab, wie sie von der Sonne empfangen. Die Temperaturen im Innern müssen daher höher als bei den erdartigen Planeten sein. Bei der Modellierung des inneren Aufbaus der beiden Planeten ergeben sich durch die unzureichenden Kenntnisse der Zustandsgleichung für Wasserstoff und für Wasserstoff-Helium-Gemische bei den zu erwartenden Drücken und Temperaturen erhebliche Unsicherheiten. Man kann daher nur grobe Charakteristika für die innere Struktur angeben.

In den Atmosphären ist jeweils eine dichte Wolkendecke vorhanden. Mit zunehmender Tiefe steigt die Temperatur. Bei Jupiter werden dabei schon ≈700 km unter der Wolkendecke ≈2 000 K erreicht. Dabei geht die Atmosphäre wahrscheinlich ohne einen die Planetenoberfläche definie-

Tab. 12.1.1–4 Die chemischen Hauptbestandteile der Atmosphären der erdartigen Planeten in den oberflächennahen Schichten (in Volumenprozent)

Bestandteile	Venus	Erde	Mars
CO_2	96,4	0,03	95,32
N_2	3,4	78,08	2,7
O_2	< 0,01	20,95	0,13
Ar	< 0,01	0,93	1,6
H_2O	≈ 0,1	0,1…2,8	0,03

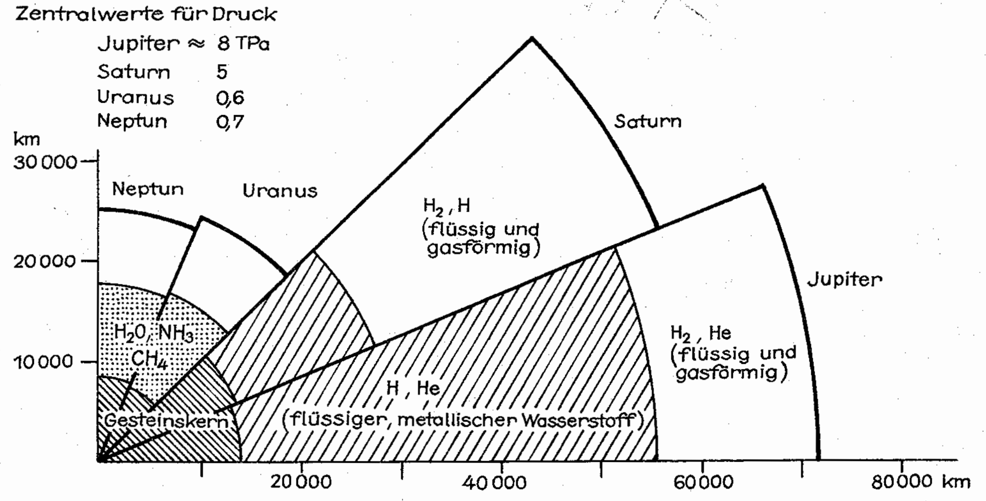


Abb. 12.1.1–6 Der Schalenaufbau der jupiterartigen Planeten

renden Phasensprung stetig in das Planeteninnere über, da die kritischen Werte für Druck und Temperatur für molekularen Wasserstoff wesentlich unter den in Jupiter vorhandenen Werten liegen. Bei einem Druck von ≈200 GPa, der in einem Zentrumsabstand von ≈0,78 Jupiterradien erreicht wird ($\varrho \approx 1\,000$ kg/m³, $T \approx 10^4$ K), geht der molekulare Wasserstoff in eine metallische Modifikation über. In diesem fluiden Zustand herrscht Konvektion, in deren Folge wahrscheinlich das Jupitermagnetfeld mittels eines Dynamoeffektes aufgebaut wird. Im Zentrum wird ein Gesteinskern vermutet, dessen Masse je nach Modell zwischen 10 und 30 Erdmassen liegt. Es ist aber unklar, ob dieser Kern fest oder flüssig ist. Die Zentraltemperatur beträgt ≈20 000 K. Abb. 12.1.1–6 zeigt ein typisches Jupitermodell.
Für Saturn wird ein ähnlicher Aufbau angenommen (vgl. Abb. 12.1.1–6). Möglicherweise ist der Gesteinskern von ≈20 Erdmassen von metallischem Wasserstoff umgeben, der mit Helium angereichert ist. Unter dem Einfluß des Schwerefeldes könnte das Helium aus den oberen Schichten in größere Tiefen absinken, wobei potentielle Energie frei wird, die dann an der Oberfläche zusätzlich zur absorbierten Sonnenenergie abgestrahlt wird. Da die Schicht aus flüssigem, metallischem Wasserstoff viel geringer als bei Jupiter ist, wirken sich Konvektionsvorgänge und damit verbundene Dynamoeffekte wesentlich weniger aus, was das schwächere Magnetfeld von Saturn erklären könnte.
Uranus, Neptun. Über den Aufbau dieser Planeten ist wenig bekannt, da die Beobachtungsgrundlagen noch sehr unsicher sind. Wegen der vergleichbaren (Uranus) bzw. etwas höheren (Neptun) mittleren Dichte bei wesentlich kleinerer Masse im Vergleich zu Jupiter kann in der Stoffzusammensetzung Wasserstoff nicht die bestimmende Rolle spielen. Vielfach nimmt man für beide Planeten ein Dreischichtenmodell an: Um einen Gesteinskern von 3 bis 4 Erdmassen befindet sich eine Schicht, in der Wasser, Ammoniak und Methan in flüssiger Form vorherrschen, während in der äußeren Zone flüssiger molekularer Wasserstoff und Helium sowie feste eisartige Beimengen, vor allem von Methan aber auch von Wasser vorhanden sind. Eine metallische Wasserstoffphase ist wegen des Stoffgemisches und des geringen Druckes wahrscheinlich nicht vorhanden.

12.1.1.5. Planetoiden. Bei den Planetoiden handelt es sich um kleine, planetenartige Himmelskörper, die in ihrer Mehrzahl nur wenige Kilometer groß sind, maximal einen Radius von 508 km (Ceres) haben. Für diejenigen, bei denen Radius und Masse unabhängig bestimmbar sind, ergeben sich mittlere Dichten zwischen 2 300 und 3 400 kg/m³. Es handelt sich daher sicherlich um Gesteinskörper, teilweise aber mit unterschiedlicher Zusammensetzung, wie Unterschiede im Reflexionsspektrum erkennen lassen. Bei einigen scheint auch Eisen eine wesentliche Rolle zu spielen. Die kleinen, unregelmäßig geformten Planetoiden sind wahrscheinlich das Ergebnis von Zusammenstößen und der teilweisen Zerstörung größerer Körper, deren Struktur etwa der der erdartigen Planeten entsprochen haben dürfte.

12.1.1.6. Planetensatelliten. Bei den Satelliten gibt es Körper, deren Größe die des Merkur übersteigt (Ganymed und Titan), neben solchen, die weitaus kleiner sind als ein mittlerer Planetoid, z. B. der Jupitersatellit Leda mit einem Radius von ≈5 km. Satelliten mit Radien ≳200 km sind kugelförmig, z. B. der Mond oder die vier großen Jupitersatelliten (Io, Europa, Ganymed, Kallisto) sowie der Saturnsatellit Titan, der als einziger eine permanente Atmosphäre hat. Die kleineren Körper sind i. allg. unregelmäßig geformt. Soweit die Dichte bestimmt werden kann, liegen die Werte bei ≈3 300 kg/m³ (Mond, Io, Europa) oder zwischen ≈1 200 und 1 900 kg/m³. Hinweise auf den inneren Aufbau erhält man wie bei den erdartigen Planeten aus Beobachtungen der Oberflächen: Sind Einschlagkrater so häufig wie bei Merkur, haben die Körper eine starre Gesteinskruste ohne magmatische Aktivitäten, sind hingegen wenig oder keine Einschlagkrater erkennbar, müssen die oberflächennahen Schichten eine plastische Struktur aufweisen, oder es herrscht starker Vulkanismus.
Die Schalenstruktur des Mondes (vgl. Abb. 12.1.1–3) ist aufgrund seismologischer Beobachtungen bekannt. Sie unterscheidet sich wahrscheinlich kaum von der des Merkur. Auch Io und Europa dürften einen eisenreichen Kern haben, der bei Europa wahrscheinlich von einem Gesteinsmantel und einer weniger als 100 km dicken, im wesentlichen aus Eis bestehenden Kruste umgeben ist. Bei Io ist der Gesteinsmantel infolge der ständig wirkenden, durch die Nachbarsatelliten Europa und Ganymed verursachten Gezeitenkräfte so stark erhitzt, daß er aufgeschmolzen ist. Die Folge ist ein starker Vulkanismus, wobei das Auswurfmaterial wesentlich aus Schwefelverbindungen besteht, so daß Io gegenwärtig eine schwefel- und schwefeldioxidreiche Kruste aufweist. Die großen Satelliten, deren mittlere Dichte unter 2 000 kg/m³ liegt, dürften etwa je zur Hälfte aus Gesteinsmaterial und aus Wasser bestehen: Ein Gesteinskern ist von einem Eis-, vielleicht auch teilweise Wassermantel umgeben, während die Kruste aus Eis, untermischt von dunklen, meteoritischen Substanzen, oder aus einem schlammähnlichen Substrat besteht (vgl. Abb. 12.1.1–7). Die kleinen Satelliten rechnet man aufgrund ihrer Reflexionseigenschaften zu den felsartigen Körpern ähnlich den Planetoiden, oder zu im wesentlichen aus Eis aufgebauten Gebilden. Die Teilchen der Saturn- und Uranusringe, die wie Satelliten die Planeten umlaufen, bestehen im wesentlichen aus Eis, doch weisen geringe Farbunterschiede möglicherweise auch auf Substanzunterschiede hin. Die Beimengung von metallischem oder silikatischem Material ist nicht auszuschließen.

12.1.2. Kometen, interplanetare Materie

12.1.2.1. Kometen bestehen aus einem Kern, der zeitweise von einer Koma und einem Schweif umgeben ist (Abb. 12.1.2–1). *Kometenkerne*, die bisher noch nicht direkt beobachtet werden konnten, haben einen Durchmesser zwischen schätzungsweise einigen 100 m und einigen Kilometern. Sie bestehen sehr wahrscheinlich aus einer innigen Verbindung von Eisbestandteilen sowie festen, meteoritischen Teilchen. Die flüchtigen Bestandteile dürften im Mittel 50 % der Gesamtmasse ausmachen, doch existieren offenbar große Unterschiede von Komet zu Komet. Die Kometen umlaufen die Sonne in meist langgestreckten Ellipsenbahnen. In Abhängigkeit von der Entfernung verdampfen unter der Wirkung der Sonnenstrahlung die gefrorenen Bestandteile (Wasser, Ammoniak, Methan, Kohlendioxid u. a.) an der Kernoberfläche und reißen dabei eingelagerte Staubteilchen mit. Gas und Staub bilden so eine zeitweilige, gravitativ nicht gebundene, vom Kern wegströmende beobachtbare Atmosphäre, die *Koma*. Die Gaskomponente ihrer inneren Schichten (bei großen, sonnennahen Kometen bis ≈10^4 km Kernabstand) besteht aus chemisch stabilen Molekülen, die in größeren Entfernungen durch die Sonneneinstrahlung teilweise zersetzt und ionisiert werden. Die meisten der beobachteten Molekülionen und Radikale

wie auch die Staubteilchen können eine Kernentfernung bis zu einigen 10^5 km erreichen; die Wasserstoffkoma aus atomarem Wasserstoff kann sich bis zu einigen 10^7 km erstrecken. Bei genügender Sonnennähe übt die Sonnenstrahlung auf die Staubteilchen (Radius einige Mikrometer) einen so starken Strahlungsdruck aus, daß sie eine zusätzliche Geschwindigkeitskomponente von der Sonne weg erhalten. In ihrer Gesamtheit bilden sie dann den gegenüber der Verbindungslinie Sonne – Komet merklich gekrümmten *Staubschweif*. Der *Plasma-* oder *Gasschweif* liegt dagegen weitgehend in dieser Linie. Er entsteht durch die Wechselwirkung zwischen den mit hoher Geschwindigkeit ($\gtrsim$400 km/s) von der Sonne abströmenden geladenen Sonnenwindteilchen und den geladenen Gaspartikeln der Koma, wobei diese auf Geschwindigkeiten von 10 bis 100 km/s beschleunigt werden. Plasmaschweife können Längen bis zu 10^8 km erlangen, ihre Dicke beträgt einige 1 000 km. Schweif und Koma sind temporäre Erscheinungen. Entfernt sich der Komet wieder von der Sonne, bilden sie sich zurück, bis sie ganz verschwinden.

12.1.2.2. Interplanetare Materie. Die aus den Kometenkernen in den interplanetaren Raum gelangenden festen Partikeln bilden die Hauptmasse des *interplanetaren Staubes*. Direkte Untersuchungen lassen sich u. a. dann durchführen, wenn Staubpartikeln auf Raumsonden treffen oder von hochfliegenden Flugzeugen aufgefangen und zur Erde gebracht werden. Ihre Struktur ist meist unregelmäßig und flockig, die Größe liegt im Bereich <1 µm bis zu einigen 10 µm. Teilchen mit Radien zwischen ≈2 bis 80 µm können durch das an ihnen gestreute Sonnenlicht (Zodiakallicht) beobachtet werden. Interplanetare Partikeln >1 mm bis zu einigen Metern bezeichnet man als *Meteorite*. Beim Eindringen in die Erdatmosphäre verursachen sie die Meteorerscheinungen. Gelangen die beim Durchqueren der Erdatmosphäre nicht vollständig verdampften größeren Teilchen bis zur Erdoberfläche, stehen sie als *Meteorite* zur direkten Laboruntersuchung zur Verfügung. Ihre Zusammensetzung variiert zwischen silikatreichem Gestein und Eisen mit einem Nickelgehalt bis zu 20 %. Die Struktur eines Teils der Meteorite zeigt, daß sie einen Differenzierungsprozeß

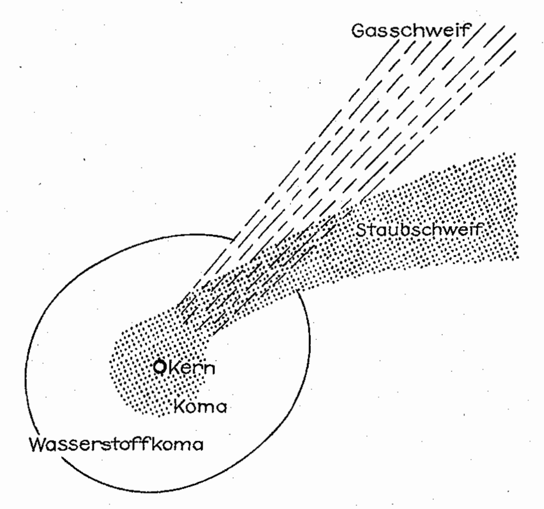


Abb. 12.1.2–1 Schema einer vollausgebildeten Kometenerscheinung

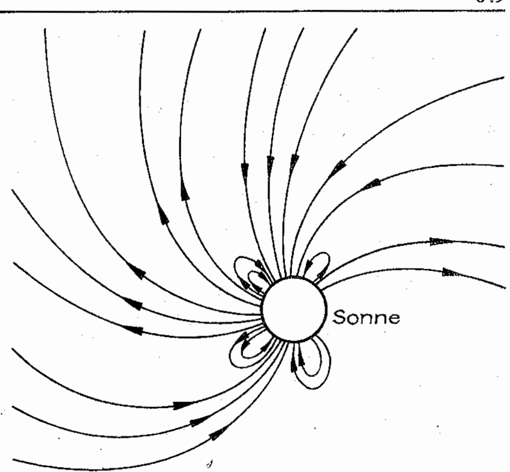


Abb. 12.1.2–2 Schematische Darstellung der magnetischen Feldlinien in Sonnennähe und im interplanetaren Raum. Der unterschiedliche Richtungssinn bewirkt eine Sektorstruktur. (Der Rotationssinn der Sonne ist in dieser Darstellung entgegen dem Uhrzeigersinn.)

innerhalb eines größeren Körpers durchgemacht haben, dessen Bruchstücke sie sind. Andere Meteorite, die Chondrite, haben hingegen noch Bestandteile, die seit ihrer Bildung unverändert geblieben sind.

Das *interplanetare Gas* ist ein heißes Plasma, das ständig von der Sonnenkorona in den interplanetaren Raum abfließt *(Sonnenwind)*. Die mittlere Strömungsgeschwindigkeit beträgt in Erdentfernung von der Sonne ≈400 bis 500 km/s, die Teilchendichte im Mittel 10 Protonen sowie 0,4 Heliumkerne je cm^3 und entsprechend viele Elektronen. Die Strömung ist großräumig turbulent. Das liegt u. a. an unterschiedlich rasch fließenden Teilströmen und an Stoßwellen, die bei Sonnenaktivitätserscheinungen entstehen. Die mittlere Plasmatemperatur beträgt wenige 10^5 K. Mit dem Plasma werden Magnetfelder mitgeschleppt, die das *interplanetare Magnetfeld* darstellen, dessen Stärke in Erdentfernung von der Sonne im Mittel $6 \cdot 10^{-9}$ T beträgt. Die Feldlinien sind infolge der Sonnenrotation spiralförmig gebogen. Das interplanetare Magnetfeld hat eine Sektorstruktur; in einem Sektor ist die Feldrichtung von der Sonne weg, im benachbarten auf sie zu gerichtet (Abb. 12.1.2–2). Beobachtungsmäßig läßt sich das interplanetare Magnetfeld mit Raumsonden direkt, aus dem Verhalten der Ionenschweife der Kometen indirekt untersuchen.

12.2. Sternphysik

Sterne sind selbstleuchtende Gaskugeln, die durch ihr eigenes Gravitationsfeld zusammengehalten werden. Bei einem Stern unterscheidet man zwei, aufgrund der Beobachtbarkeit festgelegte Gebiete: die *Sternatmosphäre*, aus der die gesamte vom Stern empfangene elektromagnetische Strahlung stammt, die damit der Beobachtung direkt zugänglich ist, und das *Sterninnere*, das mit optischen Mitteln grundsätzlich nicht beobachtbar ist. Die Sternphysik hat das Ziel, den physikalischen Zustand eines Sterns zu erschließen. Man konstruiert dazu, wie bei der Theorie des Planetenin-

nern (vgl. 12.1.1.1.), theoretische Modelle, die im wesentlichen die Druck-, Dichte- und Temperaturschichtung sowie die der chemischen Zusammensetzung der Sternmaterie beschreiben. Dabei stützen sich die Modellrechnungen in ihren numerischen Aussagen auf jeweils geeignete Beobachtungsgrößen, die als Modellparameter benutzt werden.

12.2.1. Beobachtbare Zustandsgrößen der Sterne

Beobachtungsgrößen, die einen Stern als Ganzes charakterisieren (sogenannte *Zustandsgrößen*) sind die Masse M, der Radius R und die Leuchtkraft L, das ist die gesamte Strahlungsleistung des Sterns. Aus diesen Größen läßt sich u. a. die mittlere Dichte $\bar{\varrho} = 3M/(4\pi R^3)$ und die Schwerebeschleunigung an der Sternoberfläche $g_0 = GM/R^2$ berechnen. Weiterhin ergibt sich der Strahlungsstrom an der Sternoberfläche $\Phi_0 = L/(4\pi R^2)$. Bei einem Hohlraumstrahler der Temperatur T gilt für den (über alle Wellenlängen summierten) Strahlungsstrom $\Phi_{HS} = \sigma T^4$. Man kann daher einem Stern eine *Effektivtemperatur* T_{eff} zuordnen, die der Temperatur desjenigen Hohlraumstrahlers entspricht, dessen Gesamtstrahlungsstrom gleich dem an der Sternoberfläche ist; damit erhält man $\sigma T_{eff}^4 = L/(4\pi R^2)$.

Als Beobachtungsgrößen stehen darüber hinaus solche zur Verfügung, die sich aus dem Sternspektrum ableiten lassen. Es sind dies u. a. die *Spektralklasse*, die ein Sternspektrum als Ganzes charakterisiert, sowie die Magnetfeldstärke und die Rotationsgeschwindigkeit der Sternatmosphäre. Da das beobachtete Spektrum in der Sternatmosphäre entsteht, bezieht sich die bei der Spektralanalyse aus dem Spektrum abgeleitete chemische Zusammensetzung auch nur auf die Atmosphäre. Zwischen der Effektivtemperatur und der Spektralklasse eines Sterns besteht eine eindeutige Beziehung; beide Größen sind daher nicht unabhängig voneinander.

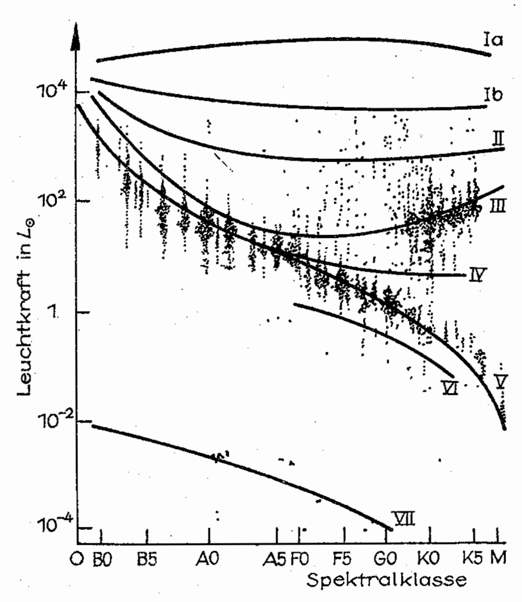


Abb. 12.2.1–1 Reales Hertzsprung-Russell-Diagramm für Sterne bis zu einer bestimmten Grenzhelligkeit. Die Linien geben die Lage der Leuchtkraftklassen an

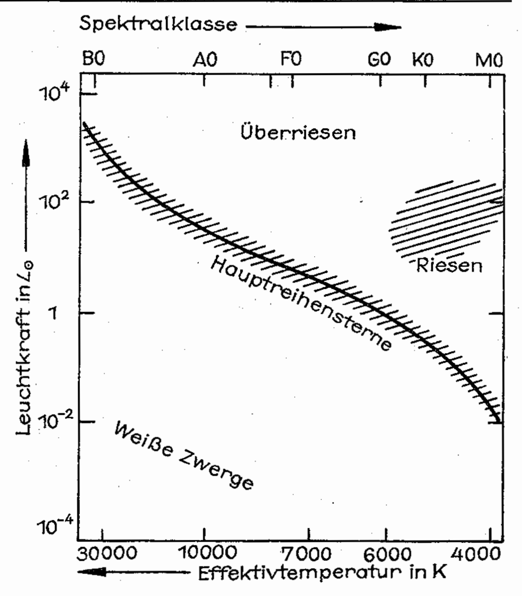


Abb. 12.2.1–2 Schematisches Hertzsprung-Russell-Diagramm

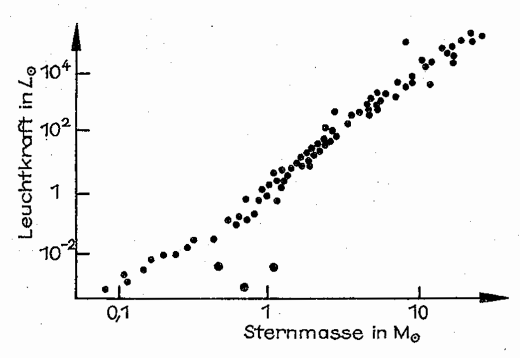


Abb. 12.2.1–3 Empirische Masse-Leuchtkraft-Beziehung. Weiße Zwerge sind mit dicken Bildpunkten gekennzeichnet

Sogenannte *Zustandsdiagramme* vermitteln grafisch den Zusammenhang zwischen zwei unabhängigen Beobachtungsgrößen. Am wichtigsten ist das *Hertzsprung-Russell-Diagramm*, Abk. *HRD*, in dem die Leuchtkraft der Sterne über der Spektralklasse bzw. der äquivalenten Effektivtemperatur aufgetragen wird (Abb 12.2.1–1). Dabei zeigt es sich, daß das HRD nicht gleichmäßig mit Bildpunkten ausgefüllt ist, sie häufen sich vielmehr in bestimmten Gebieten. Man unterscheidet entsprechend u. a. *Hauptreihensterne*, *Riesensterne*, *Überriesen* und *Weiße Zwerge* (Abb. 12.2.1–2).

Für Hauptreihensterne ergibt sich eine empirische Beziehung zwischen der Leuchtkraft und der Sternmasse, für die näherungsweise $L \sim M^{3,5}$ gilt (*Masse-Leuchtkraft-Beziehung*, Abb. 12.2.1–3). Sie wird, wie auch das HRD, im Rahmen der Theorie des inneren Aufbaus der Sterne theoretisch interpretiert.

12.2.2. Sternatmosphären

In der Sternatmosphäre ist die Dichte so gering, daß elektromagnetische Strahlung mindestens teilweise in den Weltraum gelangen kann. Das Spektrum wird dabei durch die Eigenschaften der leuchtenden Materie geprägt. Für die Theorie bildet es die wichtigste Beobachtungsgrundlage, um den *Aufbau der Sternatmosphäre*, d. h. die Druck-, Dichte- und Temperaturschichtung sowie die chemische Zusammensetzung in ihr, herzuleiten. Die Aufstellung eines exakten *Atmosphärenmodells* für einen bestimmten Stern ist aber außerordentlich kompliziert.
In der Sonnenatmosphäre unterscheidet man aufgrund der unterschiedlichen physikalischen Bedingungen drei Schichten (vgl. 12.2.2.5.). Die Photosphäre ist die unterste. Aus ihr stammt der weitaus größte Teil des beobachtbaren Lichtes; auf sie bezieht sich daher auch der größte Teil des verfügbaren Beobachtungsmaterials. Dies läßt sich auf die übrigen Sterne übertragen. Die Theorie der Sternatmosphären beschäftigt sich daher i. allg. auch nur mit den *Sternatmosphären*.
12.2.2.1. Sternspektren bestehen aus einem Kontinuum, überlagert von Absorptionslinien; in Sonderfällen sind auch Emissionslinien vorhanden. Die Intensitätsverteilung des Kontinuums ist abhängig von der Effektivtemperatur: Steigt sie, verschiebt sich das Strahlungsmaximum zu kleineren Wellenlängen hin (vgl. Wiensches Verschiebungsgesetz). Das Linienspektrum wird von der chemischen Zusammensetzung der leuchtenden Materie, aber auch ganz entscheidend von ihrem physikalischen Zustand, also der herrschenden Dichte und Temperatur, bestimmt, wobei die Dichte wiederum wesentlich eine Funktion der Schwerebeschleunigung g_0 ist.
Die Unterschiede der physikalischen Bedingungen in den Sternatmosphären spiegeln sich in der Verschiedenheit der Sternspektren wider. Von mehr als 99 % aller hellen Sterne lassen sich die Spektren phänomenologisch in ein zweidimensionales Schema mit der Spektralklasse (entsprechend der Effektivtemperatur) und der *Leuchtkraftklasse* (entsprechend der Schwerebeschleunigung) als Klassifikationsparameter einordnen. Die Hauptspektralklassen (Hauptsequenz) werden im Sinne fallender Effektivtemperatur mit den Buchstaben O, B, A, F, G, K, M bezeichnet, wobei (mit Ausnahme der Spektralklasse O, die bei 03 beginnt) jede Klasse noch dezimal unterteilt wird, indem eine der Ziffern 0 bis 9 den Buchstaben angehängt wird. Die Leuchtkraftklassen, die mit den römischen Zahlen I bis V gekennzeichnet werden, stellen innerhalb einer Spektralklasse eine Folge steigender Schwerebeschleunigung dar. Abb. 12.2.2–1 veranschaulicht die Variation einiger markanter Linien in der Spektralklassensequenz schematisch; Abb. 12.2.1–1 zeigt die Lage der Leuchtkraftklassen im HRD.
Als Maß für die Stärke einer Spektrallinie benutzt man die *Äquivalentbreite W*, die die gesamte im Bereich der Linie zurückgehaltene Strahlung angibt. Bezeichnet $S_c(\lambda)$ den empfangenen Strahlungsstrom des kontinuierlichen Spektrums in Abhängigkeit von der Wellenlänge λ und $S(\lambda)$ den Strahlungsstrom in der Linie so gilt

$$W = \int_{\text{Linie}} [S_c(\lambda) - S(\lambda)]/S_c(\lambda)\,d\lambda,$$

wobei sich die Integration über den Linienbereich erstreckt. Die Äquivalentbreite entspricht anschaulich der Breite eines rechteckigen Streifens im Spektrum mit der Höhe des Kontinuums, dessen Fläche gleich der zwischen dem Linienprofil und dem interpolierten Kontinuum ist

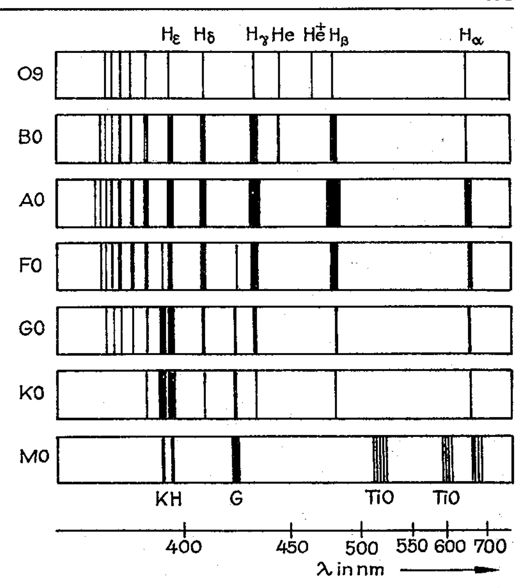


Abb. 12.2.2–1 Die Folge der Spektralklassen. Die Breite der dargestellten charakteristischen Linien ist ein Maß für die Linienstärke. H_α, H_β usw. bezeichnen die Balmer-Serie des Wasserstoffs, He bzw. He^+ Linien vom neutralen bzw. ionisierten Helium, H und K vom einmal ionisierten Calcium, TiO von dem Molekül Titaniumoxid, G das sogenannte G-Band, in dem Linien von Eisen, Titanium und Calcium dicht benachbart liegen

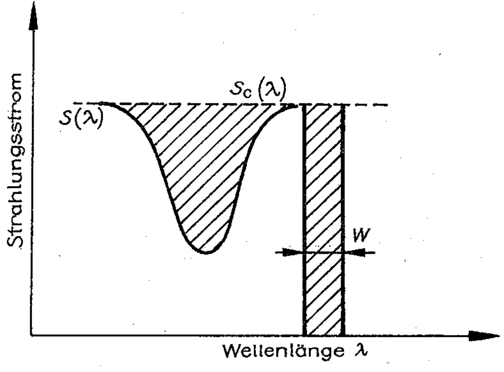


Abb. 12.2.2–2 Zur Äquivalentbreite; die schraffierten Flächen haben gleichen Inhalt

(Abb. 12.2.2–2). Bei der praktischen Bestimmung von *W* treten oft Schwierigkeiten auf, da sich Linien häufig überlappen, und daher in ihrer Umgebung vielfach kein ungestörtes Kontinuum existiert.
12.2.2.2. Thermodynamischer Zustand des Atmosphärengases. Die von der Sternatmosphäre ausgestrahlte Energie wird nicht in ihr freigesetzt, sondern gelangt aus dem Sterninnern „von unten" in sie hinein. Die Sternatmosphäre stellt nur die letzte Station bei den vielen Wechselwirkungsprozessen zwischen der Sternmaterie und der nach außen strömenden Energie dar, bevor diese in den Welt-

raum abgestrahlt wird. In der Sternatmosphäre herrscht dabei eine Temperatur- und Dichteschichtung. Das bedeutet, daß sich die Atmosphäre weder als Ganzes noch in Teilen in einem idealen thermodynamischen Gleichgewichtszustand befindet, wie er z. B. in einem Hohlraum konstanter Wandtemperatur existiert. Dennoch zeigt es sich, daß man bei den meisten Sternen jedem Volumenelement der Photosphäre einen festen Temperaturwert zuordnen kann, der sowohl die Verteilung der kinetischen Energie der Gaspartikeln (Maxwell-Verteilung) als auch die Verteilung der Atome auf die verschiedenen Anregungs- und Ionisationszustände beschreibt. Man spricht dann von einem *lokalen thermodynamischen Gleichgewicht*, so als ob jedes Volumenelement für sich ein abgeschlossenes thermodynamisches Gleichgewichtssystem bilden würde.

Bezeichnet $n(X_{l,m})$ die Anzahl der Atome eines Elementes X je Volumeneinheit, die sich im l-ten Ionisationszustand und im m-ten Anregungszustand befinden, $n(X_{l,1})$ entsprechend die Anzahl der Atome im l-ten Ionisationszustand aber im Grundzustand, so ist der *Anregungsgrad* $n(X_{l,m})/n(X_{l,1})$ im Falle des thermodynamischen Gleichgewichts allein eine Funktion der Temperatur T:

$$\frac{n(X_{l,m})}{n(X_{l,1})} = \frac{g(X_{l,m})}{g(X_{l,1})} \exp[-E(X_{l;m,1})/kT] \qquad (12.2.)$$

(Boltzmann-Gleichung). $E(X_{l;m,1})$ bedeutet dabei die Energiedifferenz zwischen dem Grundzustand und dem m-ten Anregungszustand (Anregungsenergie). $g(X_{l,m})$ und $g(X_{l,1})$ sind statistische Gewichte, die berücksichtigen, daß die betrachteten Energiezustände entartet sein können. Beschreibt weiterhin $n(X_{l+1,1})$ die Zahl der Atome des Elementes X im Grundzustand des $(l+1)$-ten Ionisationszustandes, so wird der *Ionisationsgrad* des Elementes X im thermodynamischen Gleichgewicht durch

$$\frac{n(X_{l+1,1})}{n(X_{l,1})} n_e = 2 \frac{g(X_{l+1,1})}{g(X_{l,1})} \frac{(2\pi m_e kT)^{3/2}}{h^3} \qquad (12.3.)$$
$$\exp[-E(X_{l+1;l})/kT]$$

gegeben *(Saha-Gleichung)*, wobei m_e die Elektronenmasse und $E(X_{l+1;l})$ die Ionisationsenergie bedeuten. Der Ionisationsgrad ist danach außer von der Temperatur auch noch von der herrschenden Elektronendichte n_e abhängig. Bei vorgegebener Temperatur sinkt er mit wachsendem n_e, da mehr Elektronen mit den Ionen rekombinieren können, was zur Erniedrigung des Ionisationsgrades führt.

Boltzmann- und Saha-Gleichung zusammen vermögen Einzelheiten der Sternspektren, u. a. die Spektralklassifikation als Temperaturfolge und die spektralen Unterschiede zwischen z. B. Riesen- und Zwergsternen als Ergebnis unterschiedlicher Elektronendichte (entsprechend unterschiedlicher Gesamtdichte bzw. Schwerebeschleunigung g_0), qualitativ zu erklären: Das Maximum der Stärke der Balmer-Linien bei der Spektralklasse A0 (vgl. Abb. 12.2.2–1) entsteht z. B. dadurch, daß mit wachsender Temperatur der Bruchteil der Wasserstoffatome im ersten Anregungsniveau (von dem die Balmer-Linien ausgehen) zunimmt, weil der Anregungsgrad steigt. Andererseits steigt aber auch der Ionisationsgrad, wodurch der Bruchteil der neutralen Atome relativ zur Gesamtmenge des vorhandenen Wasserstoffs fällt. Bei A0 erreicht nun gerade die Zahl der Wasserstoffatome im ersten Anregungsniveau relativ zur Gesamtwasserstoffmenge ein Maximum.

Linien, die von Ionen absorbiert werden (z. B. die He-II-Linien in Abb. 12.2.2–1), erscheinen bei höheren Effektivtemperaturen als die von neutralen Atomen des gleichen Elements, da zur Ionisation höhere Temperaturen benötigt werden. Bei sehr niedrigen Effektivtemperaturen (z. B. Spektralklasse M) beobachtet man Linien von Molekülen, die bei höheren Temperaturen nicht mehr existieren können, weil sie infolge von Dissoziation zerstört werden.

Die spektralen Unterschiede zwischen Riesen- und Überriesensternen gegenüber Hauptreihensternen erklären sich daraus, daß bei vergleichbarer Sternmasse jene einen viel größeren Radius haben, daher eine wesentlich geringere Schwerebeschleunigung g_0, und mithin geringere Atmosphärendichte aufweisen als diese. Nach der Saha-Gleichung ist demnach der Ionisationsgrad in der Atmosphäre eines Riesensterns höher als in der eines Zwergsterns gleicher Spektralklasse. Spektrallinien, die empfindlich auf die Änderung des Ionisationsgrades reagieren, können mithin als Leuchtkraftindikatoren herangezogen werden.

In den Sternatmosphären gilt die Zustandsgleichung des idealen Gases. Die überwiegend elastischen Stöße der Teilchen untereinander bewirken einen ständigen Energieaustausch, so daß unelastische Stöße, bei denen infolge von Anregung oder Ionisation dem stoßenden Partikel ein Teil seiner kinetischen Energie entzogen wird, keine merkliche Abweichung von einer Gleichgewichtsverteilung (Maxwell-Verteilung) der Geschwindigkeiten verursachen. Die große Zahl von Stößen eines Atoms je Sekunde relativ zur Zahl der Absorptions- und Emissionsprozesse, die es im gleichen Zeitraum erleidet, ist auch der Grund, weshalb man mit lokalem thermodynamischem Gleichgewicht rechnen kann. In den höchsten Atmosphäreschichten, in denen die Dichte, damit die Zahl der Stöße je Zeiteinheit, sehr gering ist, treten stärkere Abweichungen des Anregungsgrades von dem im thermodynamischen Gleichgewicht auf: Ist die Zeit zwischen zwei anregenden Stößen für ein Atom vergleichbar oder größer als die mittlere Verweilzeit im angeregten Zustand, so wird es durch Strahlungsübergänge schneller auf niedrigere Energieniveaus gelangen als umgekehrt durch Stöße auf das angeregte Niveau. Niedrige Energieniveaus sind daher gegenüber dem thermodynamischen Gleichgewichtszustand überbesetzt.

12.2.2.3. Strahlungsprozesse in Sternatmosphären. Der Energietransport erfolgt in den Sternatmosphären überwiegend durch Strahlung. Andere Energieformen (mechanische, magnetische) tragen kaum bei, obwohl sie für den Atmosphärenaufbau, speziell den der über der Photosphäre lagernden Schichten (z. B. Chromosphäre und Korona bei der Sonne, vgl. 12.2.2.5.), wichtige Konsequenzen haben können. In den Photosphären der Sterne mit Effektivtemperaturen kleiner oder gleich der der Sonne herrscht, wie in dieser, auch Konvektion. Die dabei zusätzlich zur Strahlung nach außen transportierte Energie kann aber gegenüber der durch Strahlung transportierten vernachlässigt werden; es herrscht *Strahlungsgleichgewicht*.

Beim Durchgang von Strahlung durch Materie wird ein Teil absorbiert oder gestreut, wobei man zwischen kontinuierlicher und Linienabsorption unterscheidet. Eine *kontinuierliche Absorption* erfolgt bei der Fotoionisation von Atomen oder Ionen *(gebunden-freie Übergänge)*. Dabei gelangt ein Elektron aus einem an den Atomkern gebundenen Zustand in einen freien, in dem seine kinetische Energie jeden Wert haben kann. Bei der Rekombination der Elektronen mit den Ionen wird entsprechend ein Kontinuum emittiert. Da ein freies Elektron bei elastischen Stößen einen Teil seiner kinetischen Energie mit dem Stoßpartner austauscht, hat das bei der Rekombination emittierte Photon i. allg. eine

andere Wellenlänge als das absorbierte. Bei einem nahen Vorübergang eines freien Elektrons an einem Ion kann das Elektron auch ein Photon absorbieren (oder emittieren), d. h. einen *frei-freien Übergang* ausführen, und dabei seine kinetische Energie erhöhen (bzw. erniedrigen). Langsame Elektronen haben eine höhere Absorptionswahrscheinlichkeit als schnelle, wobei längerwelliges Licht stärker absorbiert wird als kürzerwelliges. Frei-freie Übergänge spielen daher im Infrarot- und Radiowellenbereich die größte Rolle. Die *Streuung von Photonen* kann sowohl an freien Elektronen (Thomson-Streuung) als auch an Atomen und Molekülen (Rayleigh-Streuung) erfolgen. Die Wellenlänge bleibt dabei ungeändert, es tritt nur eine Richtungsänderung ein. Das emittierte Spektrum ist folglich durch die einfallende Strahlung bestimmt, nicht aber durch die lokale Temperatur wie bei den beiden ersten Prozessen, bei denen das Spektrum von der Geschwindigkeit der Elektronen, also deren (kinetischer) Temperatur abhängt.

In kühlen Sternatmosphären trägt das negative Wasserstoffion (H^-), das unter Strahlungsemission durch Anlagerung eines Elektrons an ein neutrales Wasserstoffatom entsteht, wesentlich zur Absorption bei, trotz seiner sehr geringen Häufigkeit. Das zusätzliche Elektron ist nur schwach gebunden. Licht mit einer Wellenlänge $\lambda < 1{,}655\,\mu m$ kann das Elektron wieder in den freien Zustand überführen.

Linienabsorption- bzw. *-emission* entsteht bei Übergängen zwischen diskreten Energieniveaus der Atome oder Ionen *(gebunden-gebundene Übergänge)*. Selbst für ein vollständig ungestörtes ruhendes Atom ist die einem bestimmten Übergang entsprechende Spektrallinie nicht unendlich schmal, weil der endlichen Lebensdauer eines angeregten Energieniveaus nach der Heisenbergschen Unschärferelation eine gewisse, endliche Energieunschärfe entspricht. Die Absorptionsfähigkeit eines Atoms für einen bestimmten Übergang in Abhängigkeit von der Wellenlänge wird durch den Linienabsorptionskoeffizienten beschrieben. Er hat seinen größten Wert in der Linienmitte und fällt in den Linienflügeln rasch ab. Der Abstand der beiden Punkte im Linienprofil, in denen die Intensität gleich der halben im Linienzentrum ist, bezeichnet man als *Linienbreite*. Sie hängt von verschiedenen Verbreiterungsmechanismen ab. Die geringste Verbreiterung („natürliche Linienschärfe“) gibt die genannte Energieunschärfe. Sie ist gegenüber der *Doppler-Verbreiterung* zu vernachlässigen. Bei dieser treten infolge der thermischen Bewegung der Atome Doppler-Verschiebungen der Absorptionswellenlänge auf, wobei die Häufigkeit einer bestimmten Geschwindigkeit längs der Sichtlinie (damit einer bestimmten Wellenlängenverschiebung) durch die Temperatur des Gases bestimmt wird. Mit wachsender Temperatur nimmt die Doppler-Verbreiterung zu. Neben der thermischen Bewegung können auch turbulente Strömungen über den Doppler-Effekt eine Linienverbreiterung bewirken. Ein anderer Verbreiterungsmechanismus ist die *Druckverbreiterung (Stoßdämpfung)*, die durch Stöße der absorbierenden Atome verursacht wird, weil dadurch die mittlere Verweilzeit im angeregten Zustand reduziert wird. Da die Zahl der Stöße mit der Dichte wächst, entstehen insgesamt in einer dichten, heißen Sternatmosphäre (z.B. Hauptreihensterne hoher Effektivtemperatur) breitere Spektrallinien als in einer weniger dichten und kühlen (z. B. bei kühlen Riesensternen). In der Umgebung des Linienzentrums bestimmt der Doppler-Effekt das Linienprofil, in den Linienflügeln dominiert die Stoßdämpfung.

Die Äquivalentbreite einer Spektrallinie wird durch die Zahl der in der Sichtlinie liegenden absorptionsfähigen Atome bestimmt; sie steigt mit wachsender Atomzahl.

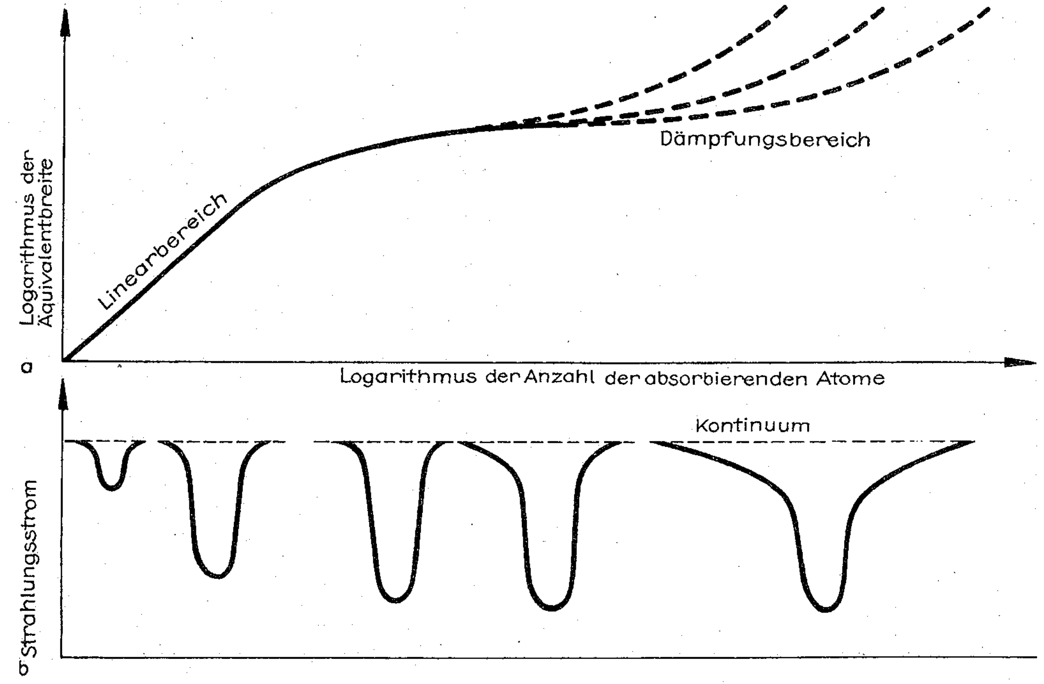


Abb. 12.2.2–3 Wachstumskurve (a); die Linienprofile (b) veranschaulichen die Entstehung der Wachstumskurve

Trägt man die Äquivalentbreite in einem Diagramm über der Zahl der absorbierenden Atome auf (Abb. 12.2.2–3), so ergibt sich eine monoton ansteigende Kurve *(Wachstumskurve)*. Ihr Anstieg ist im Fall weniger absorbierender Atome linear. Wird die Zahl größer, vergrößert sich die Wahrscheinlichkeit, daß ein absorptionsfähiges Atom gar nicht absorbieren kann, weil der Strahlungsstrom im Bereich des Linienzentrums durch andere Atome schon stark geschwächt ist. Die Äquivalentbreite wächst daher nur noch langsam, die Wachstumskurve verläuft sehr flach. Ist die Atomzahl schließlich so hoch, daß auch in den Linienflügeln eine wesentliche Absorption erfolgt, nimmt die Äquivalentbreite wieder etwas stärker zu. Die genaue Lage dieses Dämpfungsabschnittes ist von der Stärke der Stoßdämpfung abhängig. Wachstumskurven werden bei der *quantitativen Spektralanalyse* benötigt, bei der aus den beobachteten Äquivalentbreiten auf die Häufigkeit der die Linien verursachenden Elemente geschlossen wird.

12.2.2.4. Sternatmosphärenmodelle. Die geometrische Dicke einer Sternatmosphäre (Photosphäre) ist klein gegenüber dem Sternradius. Man kann daher mit einer planparallelen Schicht rechnen. Der physikalische Zustand (Druck, Temperatur, Dichte) ist dann eine Funktion der geometrischen Tiefe z. Die chemische Zusammensetzung wird für die gesamte Atmosphäre als konstant angenommen. Da die Materie für Strahlung unterschiedlicher Wellenlänge unterschiedlich durchlässig ist, stammt das beobachtete Spektrum aus ganz unterschiedlichen Tiefen: Die Absorptionsfähigkeit im Zentrum starker Linien ist z. B. viel größer als im benachbarten Kontinuum. Das beobachtete Restlicht in einer solchen Linie kann daher nur aus viel höheren Schichten als das benachbarte Kontinuum kommen. Anders ausgedrückt: Im Licht einer starken Absorptionslinie kann man weniger tief in die Atmosphäre hineinsehen als im benachbarten Koninuumslicht. Es ist daher vielfach zweckmäßig, statt der geometrischen Tiefe eine optische Tiefe $\tau(\lambda) = n\varkappa(\lambda)z$ einzuführen. Dabei ist n die Zahl der Atome je Volumeneinheit, die bei der Wellenlänge λ absorbieren, und $\varkappa(\lambda)$ der entsprechende Absorptionskoeffizient. Ganz grob kann man sagen, daß der Hauptteil der Strahlung aus Gebieten mit einer optischen Tiefe von der Größenordnung Eins stammt. Ein großes $\varkappa(\lambda)$ bedingt daher bei festem n ein kleines z.

Das Ziel der Theorie der Sternatmosphären ist es, für einen gegebenen Stern mit bekannten Zustandsgrößen und bekanntem Spektrum ein Atmosphärenmodell herzuleiten, d. h., die Druck-, Dichte- und Temperaturschichtung sowie die chemische Zusammensetzung des Atmosphärengases zu bestimmen. Man stößt dabei auf die prinzipielle Schwierigkeit, daß aus dem beobachteten Spektrum erst dann der physikalische Zustand hergeleitet werden kann, wenn die chemische Zusammensetzung bekannt ist, weil von dieser die Wechselwirkung der Strahlung mit der Materie, somit Druck und Temperatur mit bedingt wird. Umgekehrt läßt sich die Zusammensetzung erst dann berechnen, wenn man weiß, unter welchen physikalischen Bedingungen sich die absorbierende Materie befindet, da diese die Absorptionsbedingungen festlegen. Eine geschlossene Lösung des Problems ist daher nicht möglich. Man muß auf Näherungsverfahren zurückgreifen, die umständlich und langwierig sind.

Im ersten Schritt führt man eine *Grobanalyse* durch. Man nimmt hierfür an, daß die Sternatmosphäre homogen aufgebaut ist, daß also in ihr Druck und Temperatur konstant sind. Geeignete Mittelwerte für diese Größen leitet man aus den beobachteten Zustandsgrößen der Sterne ab. Macht man die Voraussetzung, daß die Linien aus dem von unten in die Atmosphäre eingestrahlten Kontinuum herausabsorbiert werden, lassen sich aus den gemessenen Äquivalentbreiten mittels Wachstumskurven Näherungswerte für die chemische Zusammensetzung bestimmen.

Als nächstes kann man eine sogenannte *graue Atmosphäre* berechnen, in die als Modellparameter die Effektivtemperatur, die Schwerebeschleunigung sowie die chemische Zusammensetzung aus der Grobanalyse eingehen. Wesentlich ist dabei die Annahme eines wellenlängenunabhängigen Absorptionskoeffizienten $\bar{\varkappa}$, der aber noch temperatur- und dichteabhängig sein darf. Setzt man weiter Strahlungsgleichgewicht voraus, läßt sich die Temperaturschichtung in Abhängigkeit von der optischen Tiefe τ in einfacher Weise angeben, wobei T_{eff} als Parameter fungiert. Der Verlauf des Gasdruckes, ebenfalls in Abhängigkeit von τ, errechnet sich unter der Annahme hydrostatischen Gleichgewichts aus Gleichung 12.1., wobei Zentrifugalkräfte nicht berücksichtigt werden und $g(z) = g_0$ gilt. Mit der Temperatur- und Druckschichtung ist auch die Dichteschichtung in ihrer Abhängigkeit von τ bekannt. Der Übergang zu geometrischen Tiefen erfolgt mittels $\tau = n\bar{\varkappa}z$, indem man aus der gegebenen chemischen Zusammensetzung für alle τ-Werte (für die ja Temperatur und Dichte bekannt sind) einen über alle Wellenlängen geeignet gemittelten Absorptionskoeffizienten $\bar{\varkappa}$ berechnet.

In der abschließenden *Feinanalyse* des Sternspektrums bringt man an den für jede Schicht berechneten Werten für Dichte und Temperatur so lange Verbesserungen an, bis sich der Gesamtenergiestrom, der durch Strahlung und Konvektion aufrechterhalten wird, überall in der Atmosphäre als konstant ergibt. Rechnet man danach aus, welche Linienstärken sich bei dem neuen Atmosphärenmodell aber der alten Zusammensetzung ergeben, wird man Abweichungen zwischen den beobachteten und den berechneten Werten erhalten. Durch Verbesserungen in der Elementenhäufigkeit und folgender erneuter Durchrechnung

Tab. 12.2.2–4 Der Logarithmus der Elementenhäufigkeit einiger ausgewählter Sterne (bezogen auf lg n(H) = 12,0)

Element	Sonne G2 V	τ Sco BO V	ε Vir G8 III	HD 140283 GO V
1 H	12,0	12,0	12,0	12,0
2 He	10,8	11,0		
6 C	8,6	8,1	8,5	6,3
7 N	7,9	8,3		
8 O	8,8	8,7		
10 Ne	7,6	8,6		
11 Na	6,3		6,6	3,7
12 Mg	7,6	7,5	7,5	5,2
13 Al	6,5	6,2	6,5	3,8
14 Si	7,6	7,6	7,7	5,3
16 S	7,2	7,2	7,3	
20 Ca	6,3		6,5	4,0
22 Ti	5,0		4,6	2,3
23 V	4,0		4,1	
24 Cr	5,7		5,9	3,5
25 Mn	5,4		5,5	3,4
26 Fe	7,5	7,3	7,6	5,1
28 Ni	6,3		6,4	3,8
38 Sr	2,9		2,6	0,6

werden die Differenzen schrittweise beseitigt. Das Verfahren ist mit hohem Rechenaufwand verbunden, vor allem, wenn man auch solche Spektrallinien berücksichtigt, für die die Annahme lokalen thermodynamischen Gleichgewichts nicht gerechtfertigt ist.
Einige Ergebnisse der quantitativen Spektralanalyse gibt Tab. 12.2.2–4. Dabei ist zu beachten, daß selbst in den besten Fällen die Elementenhäufigkeit nur auf etwa ±10 % i. allg. nur bis auf ±40 % genau sind. Der Stern τ Scorpii (τ Sco) ist ein Beispiel eines Hauptreihensterns hoher Effektivtemperatur, ε Virginis (ε Vir) das eines Riesensterns, während HD 140283 ein Vertreter der sogenannten Population II ist, die sich durch relative Armut an schweren Elementen auszeichnet. Außer dem Sternnamen ist auch die Spektral- und Leuchtkraftklasse angegeben.
Insgesamt zeigt es sich, daß das Häufigkeitsverhältnis von Wasserstoff- zu Heliumatomen bei allen Sternen etwa 10:1 beträgt, und daß die relative Häufigkeit der Elemente mit einer Ordnungszahl ≧6 untereinander praktisch immer gleich ist; hingegen variiert die Häufigkeit dieser Elemente bezogen auf Wasserstoff (bzw. Helium), die sogenannte *Metallhäufigkeit*. Für die meisten Sterne ist die Elementenhäufigkeit, damit auch der Metallgehalt, ähnlich der, wie sie bei der Sonne bestimmt wurde *(mittlere kosmische Elementenhäufigkeit)*, doch kennt man auch Sterne, mit einem doppelt so hohen Metallgehalt, andererseits auch solche, bei denen es nur ≈1/500 des Sonnenwertes beträgt. Diese Unterschiede sind durch die Zusammensetzung der interstellaren Materie bei der Entstehung der einzelnen Sterne bedingt. Im Laufe der Entwicklung des Milchstraßensystems reicherte sich die interstellare Materie immer mehr mit schweren Elementen an (vgl. 12.2.3.7.), so daß letztendlich die Metallhäufigkeit ein Maß für das Alter eines Sterns ist. Metallarme Sterne sind älter als metallreiche.
Bei einigen Sternen existieren anomale Häufigkeiten bestimmter Elemente. Die Ursachen dafür sind gegenwärtig vielfach noch strittig.

12.2.2.5. Sonnenatmosphäre. Die Sonne nimmt eine Sonderstellung unter den Sternen ein, da man den Strahlungsstrom, der von einzelnen Punkten der Sonnenscheibe ausgeht, messen kann, wobei sich eine Abnahme vom Zentrum zum Rand hin *(Randverdunklung)* zeigt (Abb. 12.2.2–5). Dies ist eine Folge der Temperaturschichtung in der Photosphäre. Umgekehrt kann man aus der Randverdunklung die Temperaturschichtung weitgehend direkt ableiten. (Bei allen anderen Sternen beobachtet man nur den über die ganze Sternscheibe gemittelten Strahlungsstrom.)
Tab. 12.2.2–6 gibt ein so gewonnenes Modell der Sonnenphotosphäre einschließlich der unteren Chromosphäre. τ_{500} bedeutet die optische Tiefe bei $\lambda = 500$ nm. Als Nullpunkt der geometrischen Tiefe wurde willkürlich $\tau_{500} = 1{,}00$ gewählt.
Die Sonnenscheibe läßt eine körnige Struktur *(Granulation)* erkennen. Helle Bereiche (Durchmesser 750 bis 1 500 km, Lebensdauer im Mittel 8 min) wechseln mit dunklen ab. In den hellen Granula herrscht eine Aufwärtsbewegung von ≈2 km/s, in den intergranularen Bereichen eine entsprechende Abwärtsbewegung. Die Temperaturdifferenz beträgt ≈200 K. Die Granulation ist der sichtbare Beweis, daß in der Sonnenphotosphäre und den darunterliegenden Schichten Konvektion herrscht. Ursache für diese sogenannte *Wasserstoffkonvektionszone* ist der hohe Absorptionskoeffizient in denjenigen tiefliegenden Schichten, in denen der Wasserstoff, das häufigste Element, teilweise ionisiert ist.
Über der Photosphäre liegt die *Chromosphäre*, die mit ≈10 000 km Höhe viel ausgedehnter als die Photosphäre mit ≈500 km Dicke ist. Ihre Ausstrahlung ist jedoch wegen der geringen Dichte viel niedriger als die der Photosphäre, so daß man sie nur bei totalen Sonnenfinsternissen oder mit speziellen Instrumenten beobachten kann. Die Chromosphäre fasert sich nach oben in zahlreiche Lichtzungen (Spicula) auf, deren Lebensdauer ≈10 min beträgt. Offenbar setzt sich die konvektive Gasströmung der Wasserstoffkonvektionszone nach oben fort. Da die Dichte nach außen hin abnimmt, können die bei der turbulenten Bewegung entstehenden Druckwellen in der Chromosphäre Über-

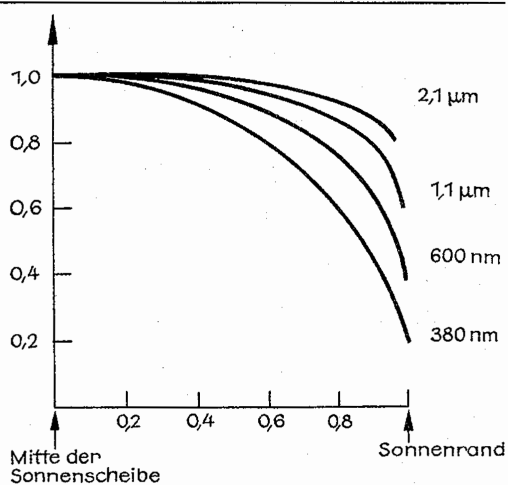


Abb. 12.2.2–5 Randverdunklung der Sonnenscheibe. Für verschiedene Wellenlängen ist das Verhältnis des gemessenen Strahlungsstromes zu dem in der Scheibenmitte in Abhängigkeit vom Zentrumsabstand aufgetragen. Das aus höherliegenden Schichten der Sonnenatmosphäre stammende längerwellige Licht ist zum Rande zu weniger geschwächt als das kürzerwellige

Tab. 12.2.2–6 Modell der Sonnenphotosphäre einschließlich der unteren Chromosphäre

τ_{500}	z in km	T in K	p in Pa	ϱ in kg/m³
$2{,}2 \cdot 10^{-5}$	−800	5 360	10,8	$3{,}09 \cdot 10^{-7}$
$3{,}53 \cdot 10^{-5}$	−700	4 890	24,1	$7{,}59 \cdot 10^{-7}$
$7{,}35 \cdot 10^{-5}$	−600	4 350	59,4	$2{,}10 \cdot 10^{-6}$
$1{,}45 \cdot 10^{-4}$	−550	4 170	96,9	$3{,}58 \cdot 10^{-6}$
$2{,}19 \cdot 10^{-4}$	−525	4 150	124	$4{,}61 \cdot 10^{-6}$
$3{,}36 \cdot 10^{-4}$	−500	4 150	160	$5{,}93 \cdot 10^{-6}$
$8{,}08 \cdot 10^{-4}$	−450	4 200	263	$9{,}63 \cdot 10^{-6}$
$1{,}92 \cdot 10^{-3}$	−400	4 330	428	$1{,}52 \cdot 10^{-5}$
$1{,}01 \cdot 10^{-2}$	−300	4 600	1 090	$3{,}64 \cdot 10^{-5}$
$4{,}79 \cdot 10^{-2}$	−200	4 920	2 610	$8{,}17 \cdot 10^{-5}$
0,100	−150	5 120	3 950	$1{,}19 \cdot 10^{-4}$
0,204	−100	5 445	5 870	$1{,}66 \cdot 10^{-4}$
0,693	−20	6 153	10 500	$2{,}63 \cdot 10^{-4}$
1,00	0	6 423	12 000	$2{,}87 \cdot 10^{-4}$
1,38	15	6 690	13 200	$3{,}04 \cdot 10^{-4}$
2,00	30	7 040	14 500	$3{,}16 \cdot 10^{-4}$
3,07	45	7 460	15 800	$3{,}25 \cdot 10^{-4}$

schallgeschwindigkeit erreichen und sich zu Stoßwellen aufsteilen. Ihre Energie geben sie schließlich an das Gas in Form von Wärme ab, was eine Temperaturzunahme mit steigender Höhe verursacht. Die Chromosphärenobergrenze liegt dort, wo Wasserstoff und Helium vollständig ionisiert sind.

Die *Korona*, die äußerste Schicht der Sonnenatmosphäre, hat eine i. allg. sehr inhomogene Struktur; entsprechend variiert die (kinetische) Temperatur des Koronagases zwischen $5 \cdot 10^5$ K („Koronalöcher") und $2 \cdot 10^6$ K. Die Dichte ist gering und nimmt nach außen hin stark ab. Der physikalische Zustand der Korona ist dabei sehr weit vom thermodynamischen Gleichgewicht entfernt. Neben einem kontinuierlichen Spektrum sendet das Gas eine Reihe von Emissionslinien aus, die u. a. von verbotenen Übergängen hochionisierter Metalle (z. B. Fe^{13+} Ca^{14+}, Fe^{9+}) herrühren, oder solche, die im Röntgen-Bereich liegen. Die hohen Temperaturen entstehen möglicherweise durch die Umwandlung mechanischer Energie (Druckwellen) in Wärmeenergie vermittels Stoßwellen (wie in der Chromosphäre), möglicherweise aber auch durch die Umwandlung magnetischer Energie in thermische mittels elektrischer Ströme (Ohmsche Wärme) oder beim Übergang komplexer Magnetfeldstrukturen in einfachere mit niedrigerer Energie. Von der Korona fließt wegen der hohen Temperaturen ständig ein Plasmastrom (Sonnenwind) in radialer Richtung ab. Er ist die Quelle des interplanetaren Gases (vgl. 12.1.2.3.).

Durch lokale Magnetfelder, deren Stärke zum Teil 0,1 T beträgt, können zeitlich und räumlich begrenzte Störungen in der Sonnenatmosphäre auftreten, die mit Sonnenflecken, -eruptionen, -fackeln usw. verbunden sind. Die Enstehung dieser lokalen Felder, die mit einer Häufigkeitsperiode von ≈ 22 Jahren auftreten, ist noch nicht restlos geklärt.

12.2.3. Innerer Aufbau der Sterne

Das Sterninnere ist per definitionem mit herkömmlichen Mitteln, d. h. mittels elektromagnetischer Strahlung, nicht beobachtbar, obwohl es praktisch die gesamte Sternmasse umfaßt. (Die Sonnenatmosphäre enthält nur etwa ein Zehnmilliardstel der Sonnenmasse!) Es geben daher nur theoretische Untersuchungen Aufschluß über die innere Struktur eines Sterns, d. h., wie Temperatur T, Dichte ϱ, Druck p und chemische Zusammensetzung χ der Sternmaterie mit dem Zentrumsabstand r variieren. (In dem Symbol χ sind die relativen Häufigkeiten aller Elemente, die in der Sternmaterie vorhanden sind, formal zu einer Größe zusammengefaßt.) Eine Überprüfung der theoretischen Ergebnisse, die in *Sternmodellen* ihren Niederschlag finden, kann daher auch nur beim indirekten Vergleich mit geeigneten Beobachtungen erfolgen.

Wegen der geringen Ausdehnung der Sternatmosphären, speziell der Photosphäre, im Vergleich zum Sternradius, kann sie für die Zwecke des inneren Aufbaus als verschwindend klein, und somit auch als *„Sternoberfläche"*, betrachtet werden.

12.2.3.1. Grundprinzipien des inneren Aufbaus. Zur einfachen theoretischen Beschreibung geht man von der für die überwiegende Mehrheit der Sterne zutreffenden Annahme aus, daß sie kugelsymmetrisch aufgebaut sind, daß sie also weder eine rasche Rotation noch starke innere Magnetfelder oder nahe Begleiter haben. Alle Größen sind dann nur von r abhängig. Weiter wird angenommen, daß sich der Stern in einem stationären Gleichgewicht befindet, daß er also z. B. keine Kontraktion oder keine Schwingungen ausführt. Es herrscht dann überall im Innern hydrostatisches Gleichgewicht, was bedeutet, daß an jeder Stelle der Druck gerade so groß ist, daß er dem Gewicht der darüberliegenden Sternschichten die Waage hält. Mathematisch wird das hydrostatische Gleichgewicht durch Formel 12.1. beschrieben, wobei jetzt $\omega = 0$ angenommenn ist, so daß der Zentrifugalterm auf der rechten Seite wegfällt. M_r ist wieder die Masse, die sich innerhalb der Kugel mit dem Radius r befindet. Geht man zur Kugel mit dem Radius $r + \mathrm{d}r$ über, so ändert sich M_r um $\mathrm{d}M_r$; der Massenerhaltungssatz liefert $\mathrm{d}M_r/\mathrm{d}r = 4\pi r^2 \varrho(r)$.

Das hydrostatische Gleichgewicht ist sehr gut realisiert. Bei einer Störung würde nämlich ein Körper außerordentlich rasch durch Kontraktion oder Expansion reagieren. Die sogenannte dynamische oder *Frei-Fall-Zeitskala* τ_{dyn}, die die Zeitspanne angibt, in der er unter dem Einfluß der Eigengravitation (bzw. eines Überdruckes) eine wesentliche Radiusänderung erleidet, ist gegenüber der Dauer anderer Entwicklungsprozesse der Sterne extrem klein. Für die Sonne ergibt sich z. B. $\tau_{\mathrm{dyn}_\odot} \approx 40$ min.

12.2.3.2. Zustandsgleichungen. Der wirksame Druck im Sterninnern setzt sich i. allg. aus dem Gasdruck p_g und dem Strahlungsdruck p_s zusammen. Da die Sternmaterie extrem weit von einer Kondensation entfernt ist, kann man sie als ideales Gas beschreiben; es gilt $p_g \sim \varrho T$. Man hat damit eine viel einfachere Zustandsgleichung als etwa für das Planeteninnere (vgl. 12.1.1.1.). Die ideale Gasgleichung gilt im Stern selbst dann, wenn die Gasdichte für irdische Verhältnisse extrem hoch, z. B. von der Größenordnung 10^5 kg/m^3 und darüber ist, da auch die Temperaturen entsprechend hoch und alle Elemente vollständig ionisiert sind. Das Eigenvolumen der freien Partikeln (Atomkerne, freie Elektronen) ist dann nämlich gegenüber dem, das einem Teilchen bei seiner Bewegung zur Verfügung steht, vernachlässigbar klein.

Es zeigt sich, daß die Strahlung im Sterninnern hochgradig isotrop ist, so daß man sie lokal als Hohlraumstrahlung ansehen kann. Der *Strahlungsdruck* ist dabei allein eine Funktion der Temperatur mit $p_s \sim T^4$. Meist ist der Strahlungsdruck wesentlich kleiner als der Gasdruck; nur im Innern massereicher Sterne leistet er einen merklichen Beitrag zum Gesamtdruck.

Bei sehr hoher Dichte (Größenordnung 10^9 kg/m^3 und höher) und vergleichsweise niedrigen Temperaturen läßt sich das aus den freien Elektronen bestehende Elektronengas nicht mehr mit der idealen Gasgleichung beschreiben, da quantenmechanische Effekte wirksam werden: Das Elektronengas entartet. Bei einem vollständig *entarteten Elektronengas* ergibt sich wieder eine relativ einfache Zustandsgleichung, da der Druck p_e der entarteten Elektronen allein von der Dichte, nicht aber von der Temperatur abhängt. Zu unterscheiden ist dabei noch zwischen einem nicht relativistisch vollständig entarteten Elektronengas mit $p_e \sim \varrho^{5/3}$ und einem relativistisch vollständig entarteten mit $p_e \sim \varrho^{4/3}$. Relativistische Entartung ist dann gegeben, wenn die kinetische Energie der Elektronen vergleichbar oder größer als ihre Ruheenergie $m_e c^2$ ist. Im Zwischenbereich zwischen normal idealem und vollständig entarteten Elektronengas muß man mit teilweiser Entartung rechnen, was zu einer komplizierteren Zustandsgleichung führt, in der der Druck temperaturabhängig ist.

12.2.3.3. Energiefreisetzungsprozesse. Befindet sich der Stern im strengen hydrostatischen Gleichgewicht, so wird

die gesamte im Innern freigesetzte Energie an der Oberfläche abgestrahlt. Bezeichnet L_r die Strahlungsleistung (Leuchtkraft) einer Kugel mit dem Radius r, so gilt für ihre Änderung dL_r beim Übergang zur Kugel mit dem Radius $r + dr$ aufgrund der Energieerhaltung $dL_r/dr = 4\pi r^2 \varrho(r) \varepsilon(r)$, wobei $\varepsilon(r)$ die spezifische Energiefreisetzungsrate infolge von Kernprozessen bedeutet.

Während bestimmter, relativ kurzer Existenzphasen eines Sternes ist das hydrostatische Gleichgewicht nicht völlig streng realisiert. Der Stern als Ganzes oder einzelne Gebiete in ihm können langsam kontrahieren oder expandieren, wobei aber immer ein quasi-hydrostatisches Gleichgewicht beibehalten wird. Dabei wird potentielle Energie frei bzw. ein Teil der inneren Energie zur Erhöhung der potentiellen Energie verbraucht. (Dies muß bei der Energiefreisetzungsrate $\varepsilon(r)$ zusätzlich berücksichtigt werden.) Die bei einer solchen Kontraktion freigesetzte potentielle Energie strahlt der Stern zur Hälfte ab. die andere Hälfte dient zur Erhöhung der inneren Energie, also der Temperatur. Dies spielt in den Entwicklungsphasen eines Sternes eine Rolle, in denen er neue Kernenergiequellen, für die eine höhere als vorhandene Temperatur benötigt wird, erschließt. Als *Helmholtz-Kelvin-Zeitskala* τ_{HK} bezeichnet man die Zeitdauer, die ein Stern mit konstanter (gegenwärtiger) Leuchtkraft strahlen könnte, falls er zur Ausstrahlung allein die bei einer Quasi-Gleichgewichtskontraktion freigesetzte potentielle Energie benützte. Generell gilt $\tau_{HK} \gg \tau_{dyn}$; für die Sonne ergibt sich z. B. $\tau_{HK\odot} \approx 10^7$ Jahre.

Die Kernenergie stellt die wichtigste Energiequelle der Sterne dar. Dabei werden aus leichten Atomkernen schwerere gebildet, deren Masse aber kleiner als die Massensumme der Ausgangskerne ist. Das Energieäquivalent Δmc^2 dieser Massendifferenz Δm *(Massendefekt)* steht dem Stern zur Ausstrahlung zur Verfügung. Die Reaktionen können nur bei hohen Temperaturen (mindestens einige 10^6 K) ablaufen. Aber selbst bei diesen haben nur wenige Kerne genügend hohe kinetische Energie, um unter Ausnutzung des Tunneleffekts so dicht aneinander heranzukommen, daß eine Kernreaktion möglich ist. Für alle anderen sind die elektrostatischen Abstoßungskräfte zwischen den positiv geladenen Kernen unüberwindbar groß. (Je größer die Ladung der beteiligten Partner ist, um so höhere Temperaturen sind erforderlich.) Der Aufbau massereicher Atomkerne aus massearmen ist nur bis Eisen ^{56}Fe möglich, da bei der Bildung noch schwererer Kerne Energie verbraucht, nicht freigesetzt wird. (^{56}Fe ist der Kern mit der höchsten Bindungsenergie.) Energetisch am günstigsten ist der Aufbau von Heliumkernen aus vier Wasserstoffkernen (Protonen), bei dem der relative Massendefekt 0,71 % beträgt. Alle anderen Prozesse sind wesentlich unergiebiger; so wird beim Aufbau von ^{56}Fe aus 56 Protonen insgesamt (über Zwischenschritte) nur 0,89 % der eingesetzten Masse für den Stern als Energieäquivalent nutzbar.

Als *nukleare Zeitskala* τ_N bezeichnet man die Zeitdauer, während der ein Stern seine Energieausstrahlung durch Kernprozesse decken kann. Allgemein gilt $\tau_N \gg \tau_{HK}$. Die Sonne z. B. verbraucht in $\approx 10^{10}$ Jahren nur ≈ 10 % ihrer gesamten Wasserstoffvorräte zur Heliumbildung und damit zur Energiefreisetzung.

Die Umwandlung von Wasserstoff in Helium kann auf zwei Wegen geschehen. Bei der *Proton-Proton-Reaktion (PP-Reaktion)* (Abb. 12.2.3–1) reagieren zunächst zwei Protonen P miteinander. Unter Aussendung eines Positrons e^+ und eines Neutrinos ν_e entsteht ein Deuteriumkern ^{2}H, der durch Aufnahme eines Protons bei Aussendung eines hoch-

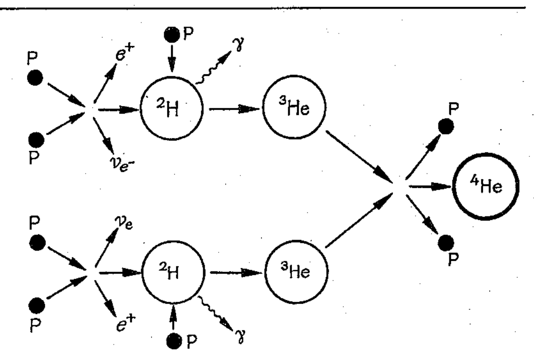


Abb. 12.2.3–1 Die Proton-Proton-Reaktion

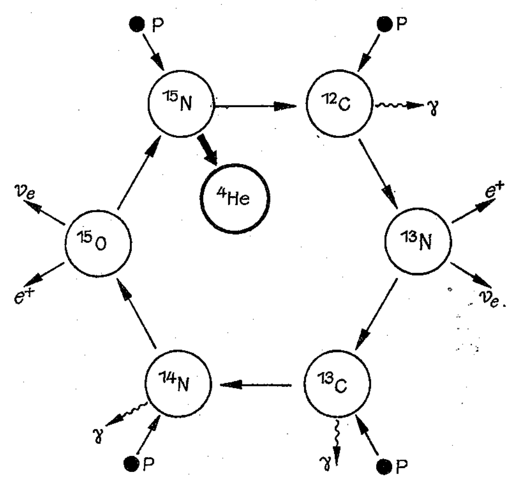


Abb. 12.2.3–2 Der Kohlenstoff-Stickstoff-Sauerstoff-Zyklus

energetischen Photons (γ-Quants) zu einem ^{3}He-Kern wird. Treffen zwei solche Kerne aufeinander, können sie unter Abgabe zweier Protonen einen ^{4}He-Kern bilden. Insgesamt werden mithin beim Aufbau eines ^{4}He-Kerns vier Protonen verbraucht. Die entstandenen Positronen zerstrahlen beim Zusammentreffen mit Elektronen infolge von Paarvernichtung, wobei die freiwerdende Energie dem Stern verbleibt. Die Neutrinos hingegen können wegen ihres extrem kleinen Wechselwirkungsquerschnittes den Stern praktisch ungehindert verlassen; sie führen dabei einen geringen Energiebetrag mit sich, der dem Stern entzogen wird. Außer dieser Hauptreaktion gibt es noch Nebenreaktionen, die jedoch recht selten stattfinden. Auch bei ihnen wird letztlich ein ^{4}He-Kern aus vier Protonen gebildet.

Beim *Kohlenstoff-Stickstoff-Sauerstoff-Zyklus (CNO-Zyklus)* (Abb. 12.2.3–2) entsteht ebenfalls ein ^{4}He-Kern aus vier Protonen. Die dabei beteiligten Kohlenstoff-, Stickstoff- und Sauerstoffkerne stehen, ohne daß sie verbraucht werden, nach Bildung von ^{4}He für weitere Prozesse zur Verfügung. Auch beim CNO-Zyklus verlassen je entstandenem ^{4}He-Kern zwei Neutrinos den Stern. Da bei diesem Prozeß Atomkerne mit höherer Kernladung beteiligt sind, werden etwas höhere Temperaturen als bei der PP-Reaktion benötigt. Die außer dem Hauptzyklus noch möglichen Nebenzyklen tragen wenig zur Gesamtenergiefreisetzung bei.

Bei Temperaturen über 10^8 K können ^{4}He-Kerne (α-Teilchen) miteinander verschmelzen *(3α-Prozeß)* (Abb. 12.2.3–3), wobei zunächst ein hochgradig radioaktiver ^{8}Be-Kern gebildet wird, dessen Halbwertszeit nur $\approx 10^{-16}$ s beträgt, und der praktisch sofort wieder in die zwei Ausgangskerne zerfällt. Nur wenn er während der kurzen Lebenszeit mit einem weiteren ^{4}He-Kern reagieren kann, bildet sich unter Aussendung eines γ-Quants ein ^{12}C-Kern. Durch Anlagerung eines weiteren α-Teilchens an den ^{12}C-Kern kann schließlich auch ^{16}O entstehen. Bei noch höheren Temperaturen sind Aufbauprozesse möglich, die letztlich bis zu ^{56}Fe führen.

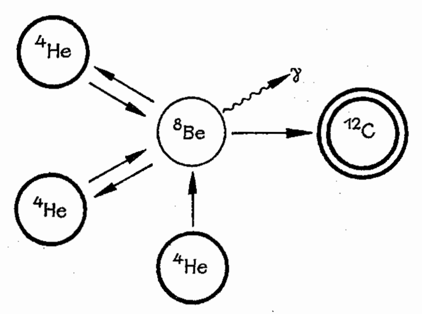


Abb. 12.2.3–3 Der 3α-Prozeß

Die spezifische Energiefreisetzungsrate ε ist abhängig von der herrschenden Temperatur, Dichte und chemischen Zusammensetzung und läßt sich mittels der von der Kernphysik experimentell oder theoretisch bestimmten Reaktionsquerschnitte sowie der bei den Reaktionen freigesetzten Energie berechnen. Es ergibt sich eine zum Teil außerordentlich starke Temperaturabhängigkeit. Es gilt z. B. bei $T \approx 10^7$ K für die PP-Reaktion $\varepsilon_{\mathrm{PP}} \sim T^5$, für den CNO-Zyklus bei $T \approx 2 \cdot 10^7$ K $\varepsilon_{\mathrm{CNO}} \sim T^{17}$ und für die 3α-Reaktion bei $T \approx 2 \cdot 10^8$ K $\varepsilon_{3\alpha} \sim T^{30}$. Die Exponenten variieren noch schwach mit T (Abb. 12.2.3–4). Wegen der starken Temperaturabhängigkeit und wegen des Temperaturanstieges zum Sternzentrum zu sind die Kernquellen dort konzentriert, und zwar um so stärker, je höher der Temperaturexponent ist.

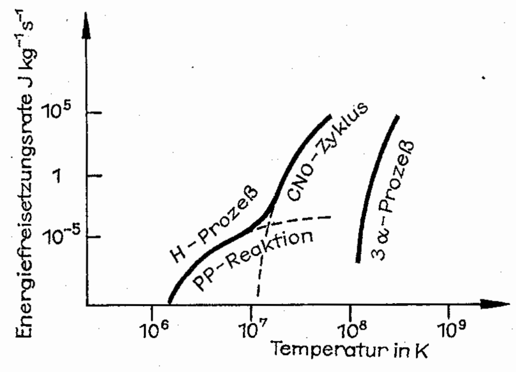


Abb. 12.2.3–4 Energiefreisetzungsrate in Abhängigkeit von der Temperatur für verschiedene Kernreaktionen. Für die Dichte und die Konzentration der an den Prozessen beteiligten Elemente sind typische Werte für die Sternzentren angesetzt

12.2.3.4. Energietransportmechanismen. Die im tiefen Sterninnern freigesetzte Energie kann nach außen durch Strahlung, Konvektion oder Wärmeleitung transportiert werden. Notwendig dazu ist aber immer ein Temperaturgefälle. *Strahlungstransport* kommt dadurch zustande, daß zwei benachbarte, unterschiedlich temperierte Gebiete sich gegenseitig Energie zustrahlen. Vom heißen (im Stern weiter innen gelegenen) Gebiet geht aber ein höherer Strahlungsstrom aus als vom kühleren (weiter außen gelegenen). Die Differenz der beiden Strahlungsströme ergibt den resultierenden, nach außen abfließenden Nettostrahlungsstrom. Für ein bestimmtes Temperaturgefälle je Längenintervall hängt die Größe des Nettostrahlungsstromes von der Durchsichtigkeit der Materie ab. Je durchsichtiger sie ist, desto größer ist die Entfernung und damit der Temperaturunterschied zwischen zwei Punkten, die sich noch direkt Strahlung zusenden können. Die Undurchsichtigkeit der Sternmaterie wird durch ihre Absorptionsfähigkeit für Strahlung bestimmt, wobei die in 12.2.2.3. genannten Absorptionsprozesse wirksam werden können. Sie sind insgesamt so effektiv, daß die freie Weglänge für Strahlung im Sterninnern im Mittel nur wenige Zentimeter beträgt, danach wird sie schon wieder absorbiert. Auf einem so kurzen Wegstück ist die Temperaturdifferenz von der Größenordnung 0,001 K. Die Strahlung ist daher tatsächlich praktisch isotrop, was bei der formelmäßigen Darstellung des Strahlungsdruckes (vgl. 12.2.3.2.) vorausgesetzt wird. Unter der Annahme von Strahlungsgleichgewicht läßt sich der Temperaturgradient $\mathrm{d}T/\mathrm{d}r$ bei Strahlungstransport berechnen. Er ist sowohl L_r als auch dem (über alle Wellenlängen geeignet gemittelten) Absorptionskoeffizienten der Sternmaterie proportional. Je höher beide Größen sind, um so größer ist $\mathrm{d}T/\mathrm{d}r$.

Bei *Konvektion* fungiert die Materie selbst als Energieträger. Heiße Materieballen (Konvektionselemente) steigen in höhere, kühlere Gebiete auf, vermischen sich dort mit der Umgebung und geben ihren Überschuß an Wärmeenergie ab. Gleichzeitig sinkt kühlere Materie ab und wird in tieferen Bereichen erwärmt. Konvektion herrscht dann, wenn ein großer Temperaturgradient vorliegt, denn dann bleiben die aufsteigenden Materieballen (trotz absoluter Abkühlung!) heißer und leichter als ihre Umgebung; sie erfahren stets neuen Auftrieb. Bei den hohen Dichten im tiefen Sterninnern kann die gesamte Energie durch Konvektion transportiert werden; es herrscht *konvektives Gleichgewicht*. Der sich dabei einstellende neue Temperaturgradient ist betragsmäßig kleiner als der bei Strahlungsgleichgewicht. Er läßt sich theoretisch aber nicht exakt berechnen, da es bisher noch keine in sich geschlossene Konvektionstheorie gibt. Das liegt vor allem daran, daß der Mischungsweg, nach dem sich die Konvektionselemente aufgelöst haben, nicht genau angebbar ist. Man kann nur Annahmen über seine Größe machen. Damit geht ein zusätzlicher freier Parameter – falls Konvektion herrscht – in das Sternmodell ein. Abschätzungen zeigen, daß der bei konvektivem Gleichgewicht sich ergebende Temperaturgradient nahe dem liegt, der sich bei einer rein adiabatischen Zustandsänderung der Materie einstellen würde (Abb. 12.2.3–5). Gebiete mit Konvektion sind wegen der guten turbulenten Durchmischung chemisch homogen.

Wärmeleitung spielt im Sterninnern nur dort eine Rolle, wo das Elektronengas entartet ist. Sie ist dann aber so effektiv, daß der Strahlungstransport dagegen vernachlässigbar wird. Es genügt ein extrem kleines Temperaturgefälle, um die gesamte Energie zu transportieren. Gebiete, in denen das

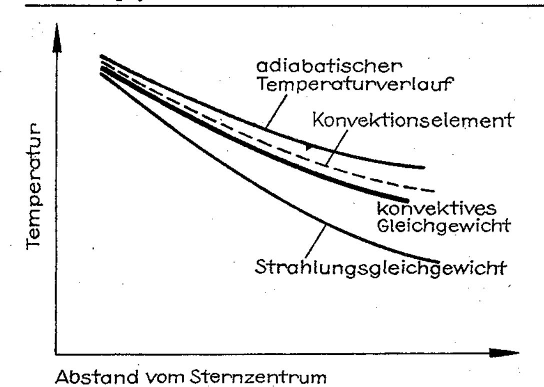


Abb. 12.2.3–5 Schematische Darstellung des Temperaturgradienten bei konvektivem Gleichgewicht im Vergleich zu dem, der in den Konvektionselementen herrscht, und dem bei rein adiabatischen Zustandsänderungen sowie dem bei Strahlungsgleichgewicht

Elektronengas vollständig entartet ist, sind daher nahezu isotherm.

12.2.3.5. Sternmodelle. Zur Berechnung des inneren Aufbaus eines Sterns, d. h. eines Sternmodells, stehen vier Differentialgleichungen zur Verfügung. Je eine beschreibt das hydrostatische Gleichgewicht, die Massenerhaltung und die Energieerhaltung, die vierte gibt den Temperaturgradienten an, wobei entweder der für das Strahlungsgleichgewicht oder der für das konvektive Gleichgewicht zu nehmen ist. Hinzu kommen noch drei algebraische Gleichungen, die die Materialeigenschaften in ihrer Abhängigkeit von T, ϱ und χ beschreibt: die Zustandsgleichung, die Energiefreisetzungsrate und der Absorptionskoeffizient der Sternmaterie. Das Gesamtgleichungssystem muß für alle Punkte im Stern gelöst werden, wobei noch zusätzliche Randbedingungen erfüllt sein müssen: Im Sternzentrum darf z. B. die Dichte weder Null noch Unendlich sein, oder an der Sternoberfläche müssen sich an die Lösung für das Sterninnere die Werte für eine Sternatmosphäre glatt anschließen lassen. Die mathematische Untersuchung des Gleichungssystems ergibt, daß eine Lösung – damit ein Sternmodell – von zwei frei wählbaren Parametern, nämlich der Sternmasse und der chemischen Zusammensetzung der Sternmaterie abhängt. (Sie beschreiben den Stern als Ganzes; der Parameter für den Mischungsweg gibt die verwendete Konvektionstheorie an.) Gibt man sich also diese beiden Größen vor, läßt sich der innere Aufbau des Sternes numerisch berechnen. Dabei zeigt es sich, daß die Lösung unter bestimmten Umständen, nämlich bei sehr inhomogener Zusammensetzung, nicht eindeutig ist. Es gibt dann zwei oder mehr Sterne mit der gleichen Masse und gleicher chemischer Zusammensetzung aber unterschiedlichem Aufbau. Wesentlich ist jedoch, daß diese Mehrfachlösungen beträchtlich voneinander abweichen, d. h., daß der innere Aufbau dieser Sterne, also Temperatur- und Dichteverteilung im Innern, sich um endliche Werte, nicht um unendlich kleine, unterscheidet. Anders ausgedrückt, in der Umgebung jeder Gleichgewichtslösung existiert weiter keine andere. Damit kann auch ein Stern von einem Zustand (einer Lösung) nicht bei unveränderter Masse und Zusammensetzung beliebig in einen anderen überwechseln, solange er sich im hydrostatischen Gleichgewicht befindet. Für instabile Phasen, bei denen das hydrostatische Gleichgewicht entscheidend gestört ist, gilt das nicht mehr.

Die Modellrechnungen liefern außer den inneren physikalischen Zustand auch die Werte für die der Beobachtung zugänglichen Größen, vor allem Leuchtkraft, Radius und Effektivtemperatur. Damit kann man im HRD die Bildpunkte der theoretischen Sternmodelle eintragen und sie mit denen von realen, beobachteten Sternen vergleichen.

12.2.3.6. Sternentwicklung. Liegt für einen Stern vorgegebener Masse und chemischer Zusammensetzung ein Sternmodell vor, so ist neben der Temperatur- und Dichteschichtung auch die Energiefreisetzungsrate ε in Abhängigkeit von r bekannt. Aus ε ergibt sich für jeden Ort im Sterninnern die Änderung der Zusammensetzung während eines Zeitintervalls Δt, da ja bei den Kernprozessen Elementumwandlungen stattfinden. Herrscht in den Energiefreisetzungszonen Strahlungsgleichgewicht, so bleiben die neu gebildeten Elemente am Ort ihrer Entstehung, bei Konvektion werden sie gleichmäßig über das gesamte Konvektionsgebiet verteilt. Mit der so geänderten chemischen Zusammensetzung läßt sich ein neues Modell berechnen, das den Zustand des gleichen Sterns aber für einen um Δt späteren Zeitpunkt seiner Existenz beschreibt. Wiederholt man das Verfahren, kann man die gesamte Sternentwicklung numerisch verfolgen. Für ihren Beginn (abgeschlossene Sternentstehung) nimmt man chemische Homogenität an, da ja zu diesem Zeitpunkt die Kernprozesse erst einsetzen, damit die chemische Veränderung. Die Häufigkeit der Elemente wird dabei kosmogonisch sinnvoll gewählt, so wie sie in der interstellaren Materie zum Zeitpunkt der Entstehung des Sterns geherrscht haben dürfte.

Für chemisch homogene Sterne wurden bisher keine Mehrfachlösungen gefunden. Damit ist die Entwicklung eines Sternes eindeutig durch seine Ausgangsmasse und -zusammensetzung bestimmt, es sei denn, er erleidet in einer späten Entwicklungsphase eine wesentliche Instabilität. Die Folge der Bildpunkte des Modellsterns im HRD, die sich aus seiner zeitlichen Entwicklung ergibt, wird als *Entwicklungsweg* bezeichnet.

Hauptreihensterne. Für chemisch homogene Sterne nimmt man i. allg. eine mittlere kosmische Elementenhäufigkeit an (vgl. 12.2.2.4.). In ihnen setzt die Umwandlung von Wasserstoff in Helium (dafür werden die geringsten Temperaturen benötigt!) gerade ein. Die Bildpunkte dieser Sterne liegen im HRD im Bereich der Hauptreihe an deren linken Grenze; sie bilden die sogenannte *Alter-Null-Hauptreihe*. Tab. 12.2.3–6 gibt die Werte für Leuchtkraft L (Einheit Sonnenleuchtkraft $L_\odot = 3{,}82 \cdot 10^{26}\,\mathrm{W}$), Zentraltemperatur T_c, Zentraldichte ϱ_c, Radius R (Einheit Sonnenradius $R_\odot = 6{,}96 \cdot 10^8\,\mathrm{m}$) und die Verweilzeit τ_{HR} auf der Hauptreihe, während der die Sterne ihre Energieausstrahlung durch die Umwandlung von Wasserstoff in Helium decken, für chemisch homogene Sterne (relativer Masseanteil von H 70 %, He 28 %, der Rest schwere Elemente entsprechend ihrer mittleren kosmischen Häufigkeit) unterschiedlicher Masse M (Einheit Sonnenmasse $M_\odot = 1{,}99 \cdot 10^{30}\,\mathrm{kg}$). In massereichen Sternen $\gtrsim 1{,}8 M_\odot$) ist der CNO-Zyklus maßgebend. Wegen seiner hohen Temperaturabhängigkeit sind die Energiequellen stark um das Sternzentrum konzentriert. Der Strahlungstransport ist dort nicht in der Lage, genügend Energie zu transportieren. Diese Gebiete sind folglich im konvektiven Gleichgewicht. Massearme Sterne, bei denen die PP-Reaktion für die Energiefreisetzung sorgt, haben dagegen (wie die Sonne, vgl. 12.2.2.5.) eine außen liegende Konvektionszone, die mit abnehmender Masse

immer mächtiger wird und unterhalb $\approx 0{,}3 M_\odot$ den gesamten Stern erfaßt. Die für die Hauptreihensterne aus Beobachtungen gefundene Masse-Leuchtkraft-Beziehung wird von den theoretischen Modellen voll reproduziert, was als Bestätigung der Richtigkeit der in die Theorie eingehenden Grundannahmen gewertet werden kann. Auch diejenigen Sterne, die nicht mehr streng homogen sind, bei denen aber

Tab. 12.2.3–6 Einige charakteristische Größen chemisch homogener Sterne (Massenanteile 70 % H, 28 % He, 2 % schwere Elemente)

M in $M_\odot$	L in $L_\odot$	T_c in K	ϱ_c in kg/m³	R om $R_\odot$	τ_{HR} in Jahren
0,25	0,01	$7{,}5 \cdot 10^6$	$155 \cdot 10^3$	0,236	$\approx 7 \cdot 10^{10}$
0,50	0,05	9,0	74	0,41	$\approx 3 \cdot 10^{10}$
1,00	0,65	14	90	0,87	$1 \cdot 10^{10}$
1,25	1,8	17	94	1,1	$4 \cdot 10^9$
1,5	4,5	19	88	1,2	$2 \cdot 10^9$
2,25	22	22	59	1,4	$5{,}3 \cdot 10^8$
3,0	79	24	40	1,8	$2{,}2 \cdot 10^8$
5,0	560	27	18	2,4	$6{,}4 \cdot 10^7$
9,0	4000	31	10	3,4	$2{,}1 \cdot 10^7$
15	20000	34	6,2	4,4	$1{,}0 \cdot 10^7$

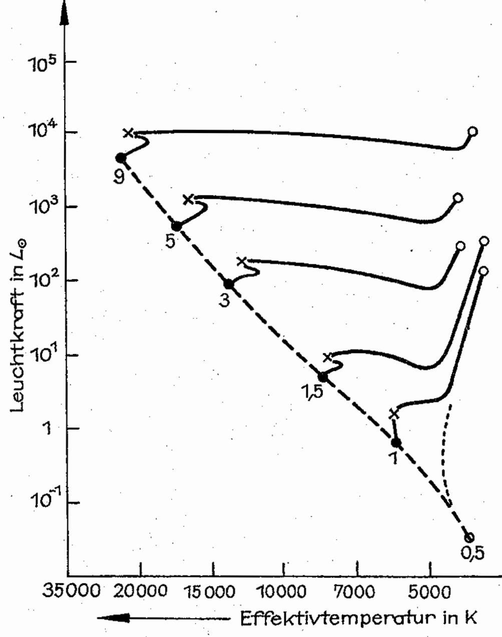


Abb. 12.2.3–7 Entwicklungswege für Sterne unterschiedlicher Masse, angegeben in Einheiten der Sonnenmasse. Die gestrichelte Linie zeigt die Alter-Null-Hauptreihe; die Kreuze markieren die Orte, wo die Wasserstoffprozesse nur noch in einer Schale um ein zentrales Heliumgebiet stattfinden, die offenen Kreise das Einsetzen der Heliumreaktion im Sternzentrum. Der Entwicklungsweg für $0{,}5\text{-}\mathfrak{M}_\odot$-Sterne ist punktiert angedeutet. Die Verweilzeit dieser Sterne auf der Hauptreihe ist länger als die Hubble-Zeit

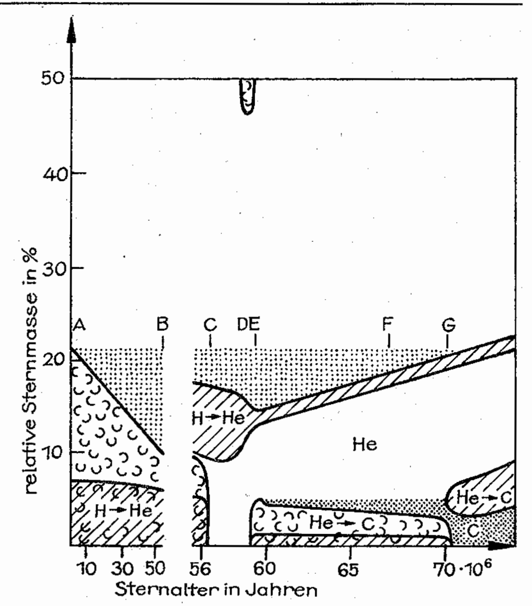


Abb. 12.2.3–8 Entwicklung des inneren Aufbaus eines Sterns von $5\mathfrak{M}_\odot$ vom Beginn des Einsetzens der Wasserstoffprozesse an. Ordinate: relative Sternmasse als Maß für den Abstand vom Sternzentrum. Abszisse: Alter des Sterns (man beachte die Streckung der Zeitskale im rechten Teil der Abb.!). In den schraffierten Gebieten ist die Energiefreisetzungsrate $\varepsilon > 10^{-1}$ W/kg, in den punktierten nimmt der H- bzw. He-Gehalt nach innen zu ab. In „wolkig" gezeichneten Gebieten herrscht Konvektion

die Energiefreisetzung durch Wasserstoffprozesse erfolgt, befinden sich noch im Gebiet der Hauptreihe. Erst bei der Erschöpfung der zentralen Wasserstoffvorräte wird diese verlassen. Die Zeit τ_{HR} ist außer von den beim Beginn der Entwicklung vorhandenen Energievorräten (die proportional der Sternmasse sind) auch von der je Zeiteinheit ausgesandten Energie, also der Leuchtkraft L, abhängig. Aufgrund der Masse-Leuchtkraft-Beziehung ergibt sich daraus näherungsweise $T_{HR} \sim M^{-2,5}$. Je massereicher ein Stern ist, um so kürzer ist seine Verweilzeit auf der Hauptreihe (vgl. Tab. 12.2.3–6), um so schneller entwickelt er sich.

Die Sonne befindet sich – obwohl vor $\approx 4{,}5 \cdot 10^9$ Jahren entstanden – noch auf der Hauptreihe. Sie ist aber nicht mehr chemisch homogen. Die zentrale Wasserstoffhäufigkeit (Massenprozent) ist von ≈ 70 % auf ≈ 50 % abgefallen, entsprechend die Heliumhäufigkeit auf knapp 50 % gestiegen. Der Anteil der schweren Elemente blieb unverändert bei ≈ 2 %. Seit der Entstehung der Sonne sind ihre Leuchtkraft und ihr Radius leicht gestiegen (vgl. Tab. 12.2.3–6), ebenso auch die Zentraltemperatur und -dichte. Diese Werte sind jedoch alle noch etwas modellabhängig (es geht z. B. die Länge des Mischungsweges ein). Mit Hilfe von Modellrechnungen für die gegenwärtige Sonne läßt sich der Neutrinostrom bestimmen, der ständig das Sonneninnere verläßt (vgl. 12.2.3.3.) und ohne Behinderung die Erde erreicht, wo er untersucht werden kann. Man erhält mittels seiner Beobachtung – im Gegensatz zu der mit elektromagnetischer Strahlung! – einen direkten Einblick in das Sonneninnere, da ja Anzahl und Energie der freigesetzten Neutrinos von

der Art und Intensität der im Sonneninnern ablaufenden Kernprozesse abhängen, und diese wiederum durch die dort herrschenden Temperatur- und Dichteverhältnisse bedingt sind. Der bei den Beobachtungen registrierte Neutrinostrom ist um einen Faktor 3 bis 5 kleiner, als man nach der Theorie erwarten sollte. Eine befriedigende Klärung dieser Diskrepanz steht noch aus. Sie könnte sowohl durch Ungenauigkeiten in der Theorie des Sonnenaufbaus als auch durch systematische Meßfehler bei den Beobachtungen oder durch noch mangelhafte Kenntnisse über das Verhalten von Neutrinos verursacht sein.

Riesensterne. Nach der Erschöpfung des zentralen Wasserstoffvorrates erfolgt die Energiefreisetzung nur noch in einer Schale um das dann hauptsächlich aus Helium bestehende Zentralgebiet. Da zur Zündung der Heliumreaktion wesentlich höhere Temperaturen erforderlich sind als für die Wasserstoffprozesse, muß das Zentralgebiet kontrahieren, denn nur durch Freisetzung von potentieller Energie läßt sich eine wesentliche Temperaturerhöhung erreichen (vgl. 12.2.3.1.). Während der zentralen Kontraktion expandieren aber die äußeren Sternregionen; im HRD durchläuft der Stern dabei einen Weg in Richtung auf das Gebiet der roten Riesensterne (Abb. 12.2.3–7). Die Dauer der zentralen Kontraktion richtet sich nach der Helmholtz-Kelvin-Zeitskale; sie ist wesentlich kürzer als die Verweilzeit auf der Hauptreihe, die durch eine Nuklearzeitskale bestimmt ist. Im Gebiet der roten Riesensterne ist die innere Struktur der Sterne komplizierter. In Zentrumsnähe wird Helium in Kohlenstoff umgewandelt, darüber liegt eine weitgehend aus Helium bestehende Schale, die von einer Schale umgeben ist, in der Wasserstoffprozesse ablaufen; außen im Stern ist noch die ursprüngliche chemische Zusammensetzung vorhanden. Abb. 12.2.3–8 zeigt die Änderung der inneren Struktur eines Sterns von 5 $M_{\odot}$ vom Beginn der Wasserstoffreaktionen bis zu einem späten Entwicklungszustand (Riesenstadium), in der der 3α-Prozeß um ein zentrales Kohlenstoffgebiet mit zur Energiefreisetzung beiträgt. Die Verweilzeit im Riesenstadium ist wieder durch eine Nuklearzeitskale bestimmt, doch ist sie wesentlich kürzer als die für die Hauptreihe, da der relative Massendefekt bei der 3α-Reaktion wesentlich kleiner als bei den Wasserstoffprozessen ist (vgl. 12.2.3.3.). In nicht zu massereichen Sternen kann im übrigen bei der Entwicklung zu einem roten Riesenstern die zentrale Dichte so hoch ansteigen, daß das Elektronengas mindestens teilweise entartet. Abb. 12.2.3–9 zeigt die Änderung von Zentraltemperatur und -dichte für den Stern von 5 $M_{\odot}$, dessen innere Entwicklung in Abb. 12.2.3–8 dargestellt ist.

Weiße Zwerge sind Sterne, in deren Innern alle Kernenergiequellen erschöpft sind, und die einen stabilen Endzustand erreicht haben. In ihnen ist die Dichte so hoch, daß das Elektronengas vollständig (nichtrelativistisch) entartet ist und praktisch allein das hydrostatische Gleichgewicht im Stern aufrechterhält. Eine Kontraktion (bei konstanter Sternmasse) ist nicht möglich, da dabei die Dichte, mithin der Elektronendruck so sehr ansteigen würde, daß der Stern mit einer sofortigen Expansion reagieren müßte. Die von Weißen Zwergen abgestrahlte Energie stammt von dem (nicht entarteten) Atomkerngas, dessen Temperatur daher ständig sinkt, während das Elektronengas wegen quantenmechanischer Bedingungen seine Energie unverändert beibehält, so daß der Sternradius nicht kleiner werden kann. Bei permanenter Ausstrahlung mit konstantem Radius sinkt die Leuchtkraft allmählich so weit ab, bis sich schließlich der Stern der visuellen Beobachtung entzieht.

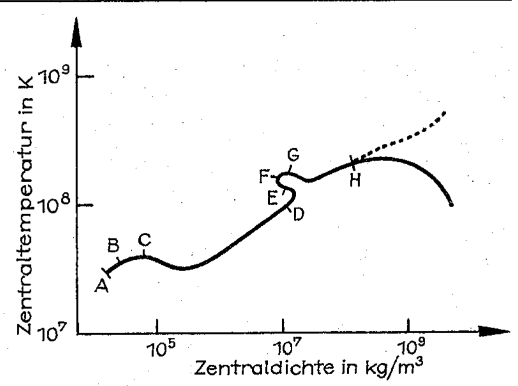


Abb. 12.2.3–9 Die Änderung der Zentraldichte und -temperatur eines Sterns von 5$\mathfrak{M}_{\odot}$ im Laufe seiner Entwicklung. Die Buchstaben kennzeichnen die gleichen Entwicklungszustände wie in Abb. 12.2.3–8. Ohne Berücksichtigung der Neutrinoemission ergäbe sich der punktierte Verlauf

Wie die Rechnungen zeigen, ist der Radius eines Weißen Zwerges von der Masse abhängig: Mit steigender Masse fällt der Radius (Abb. 12.2.3–10). Es existiert eine Grenzmasse von 1,44 $M_{\odot}$ *(Chandrasekhar-Grenzmasse)*, oberhalb der es keine Weißen Zwerge gibt (dabei ist angenommen, daß der Stern keinen Wasserstoff mehr enthält). Der Aufbau Weißer Zwerge ist relativ einfach. Die mittlere Dichte beträgt $\approx 10^9$ kg/m³ (der Sternradius ist entsprechend klein!). Wegen der hohen Wärmeleitfähigkeit des entarteten Elektronengases ist das Sterninnere nahezu isotherm. In einer äußeren dünnen Schicht ist die Materie nicht entartet. Dort fallen Temperatur und Dichte sehr stark ab. Die eigentliche Atmosphäre, von der aus die Energieabstrah-

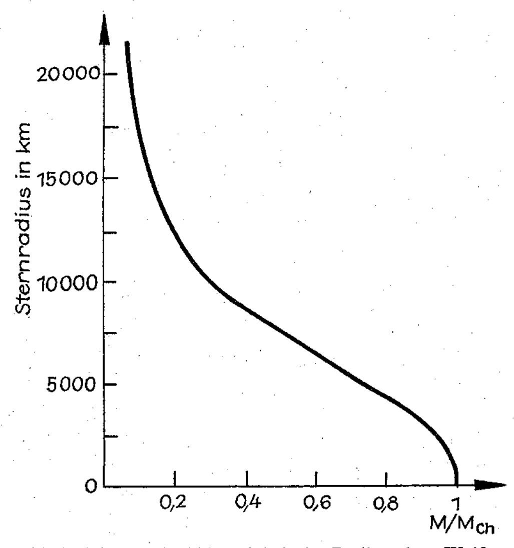


Abb. 12.2.3–10 Die Abhängigkeit des Radius eines Weißen Zwerges von seiner Masse, wobei als Einheit die Chandrasekhar-Grenzmasse $\mathfrak{M}_{\mathrm{Ch}} = 1{,}44\mathfrak{M}_{\odot}$ gewählt wurde. (Im Stern ist kein Wasserstoff mehr vorhanden.)

lung in den Weltraum erfolgt, ist größenordnungsmäßig nur ≈100 m dick.

12.2.3.7. Stabilität von Sternen. Ein Stern ist dann stabil, wenn er auf eine Störung des Gleichgewichts so reagiert, daß er wieder zum Gleichgewicht zurückkehrt; instabile Sterne entfernen sich hingegen immer mehr vom Gleichgewichtszustand. Stabilitätsuntersuchungen erfordern komplizierte theoretische Methoden. Sie erbrachten u. a. folgende Ergebnisse:

Hauptreihensterne mit Massen größer als $\approx 60\ M_\odot$ sind instabil. In ihren Zentralgebieten leistet der Strahlungsdruck mit seiner hohen Temperaturabhängigkeit einen wesentlichen Beitrag zum Gesamtdruck. Eine kleine Störung, z. B. Radiusänderung, kann sich dadurch im Laufe der Zeit zu immer größeren Schwankungen aufschaukeln, bis der Stern an seiner Oberfläche schließlich Masse abstößt. Dieser Effekt sorgt wahrscheinlich für eine obere Massengrenze stabiler Sterne, die bei $\approx 60\ M_\odot$ liegen dürfte.

Die meisten Sterne sind gegenüber kleinen Störungen in der Energiefreisetzung stabil, da ein ideales Gas bei einer kleinen zusätzlichen Energiezufuhr mit einer Expansion, und damit Abkühlung, reagiert, wodurch die Energiefreisetzung reduziert wird. Treten Kernreaktionen aber in entarteter Materie auf, z.B. die 3α-Reaktion im entarteten Zentralgebiet eines nicht sehr massereichen Sternes im Riesenstadium, so bewirkt eine Energiezufuhr eine Temperaturerhöhung, aber keine Expansion, da ja der Druck des entarteten Elektronengases nicht temperaturabhängig ist (vgl. 12.2.3.2.). Mit der Temperaturerhöhung wird die Energiefreisetzung weiter gesteigert, die Materie noch stärker aufgeheizt usw. Dies führt schließlich zu einem kurzzeitigen, exponentiellen Anstieg der Energiefreisetzung, der erst dann abgefangen wird, wenn durch eine genügend hohe Temperatur die Entartung des Elektronengases aufgehoben wird, so daß dann die Materie wie ein normales ideales Gas mit Expansion und Abkühlung reagieren kann. Im beschriebenen Fall liegt eine wesentliche Störung des Gleichgewichts vor, so daß die Eindeutigkeit des berechneten Sternmodells während dieser Instabilitätsphase nicht mehr garantiert ist (vgl. 12.2.3.5.). Wegen der hohen Undurchsichtigkeit der Sternmaterie bleibt die thermische Instabilität beim Einsetzen der 3α-Reaktion im entarteten Zentralgebiet einem äußeren Beobachter weitgehend verborgen. Beginnen hingegen Kohlenstoffumwandlungsprozesse in entarteter Materie, dann können die dabei auftretenden thermischen Instabilitäten letztlich zur Explosion des gesamten Sternes führen und eine Supernova des Typs I verursachen.

Pulsierende Sterne schwingen um eine Gleichgewichtslage. Bei ihnen wird die Schwingungsdämpfung infolge von Energiedissipation dadurch kompensiert, daß in den äußeren Sternschichten über Absorptionsprozesse in der jeweils richtigen Schwingungsphase Strahlungsenergie in mechanische Energie überführt wird. Wegen der besonderen Bedingungen, die der Absorptionskoeffizient dabei erfüllen muß, tritt eine solche Instabilität nur bei Sternen mit einer Effektivtemperatur von ≈6 000 K auf. Pulsierende Sterne können wegen ihrer periodischen Leuchtkraftänderungen als Pulsationsveränderliche beobachtet werden.

Bei *Supernovaausbrüchen* erleiden Sterne in der Spätphase ihrer Entwicklung eine extrem starke Störung ihres Gleichgewichtes. Bei massereichen Sternen kollabieren dabei die inneren Sternregionen, während die weiter außen liegenden explosionsartig weggeschleudert werden. Dabei können über Neutroneneinfang auch schwerere Elemente als Eisen aufgebaut werden, die dann mit den im Sterninnern während der vorherigen, normalen Entwicklung gebildeten Elementen in den interstellaren Raum gelangen. Die dort schon befindliche interstellare Materie wird somit im Laufe der Zeit immer mehr mit schweren Elementen angereichert. Die später aus dieser Materie gebildeten Sterne haben mithin einen höheren Metallgehalt als diejenigen, die vorher entstanden sind (vgl. 12.2.2.4.). Massearme Sterne können eventuell dann zu einer Supernova werden, wenn in ihren Zentralgebieten nach dem Verbrauch von Helium Kohlenstoffumwandlungsprozesse in entarteter Materie einsetzen. Die dabei ausgelöste thermische Instabilität führt zu einem extremen Temperaturanstieg, wodurch die Entartung des Elektronengases aufgehoben wird. In dem nunmehr wieder normal reagierenden Gas löst die Temperaturerhöhung eine explosionsartige Expansion des Sternes aus, was möglicherweise zu seiner vollständigen Zerstörung führt.

12.3. Physik der interstellaren Materie

Unter interstellarer Materie im engeren Sinn versteht man die Gesamtheit der gasförmigen und festen Teilchen, die sich im Raum zwischen den Sternen befinden. Im weiteren Sinn gehören auch das interstellare Strahlungsfeld, das interstellare Magnetfeld und die kosmische Strahlung dazu. Zwischen diesen Komponenten bestehen zum Teil enge Wechselbeziehungen.

12.3.1. Interstellares Gas

12.2.1.1. Beobachtungen. Interstellares Gas kann im optischen Spektralbereich in Emission oder in Absorption beobachtet werden; man bezeichnet es dann als leuchtend bzw. nichtleuchtend. In anderen Spektralbereichen kann aber optisch nichtleuchtendes Gas durchaus Strahlung aussenden.

Die Spektren *leuchtender Gasnebel* bestehen aus Emissionslinien neutraler und ionisierter Atome, die einem relativ schwachen Kontinuum überlagert sind. Die ausgestrahlte Energie ist umgesetzte Sternstrahlung, die von benachbarten Sternen hoher Effektivtemperatur ($T_{\rm eff} > 25\,000$ K) stammt. Durch sie wird das Gas zum Leuchten angeregt und dadurch beobachtbar. Da der Wasserstoff als häufigstes Element im interstellaren Raum in den *Emissionsnebeln* vollständig ionisiert ist, werden diese auch als *HII-Gebiete* bezeichnet.

Das im optischen Spektralbereich nicht leuchtende Gas macht sich durch Absorptionslinien in den Spektren der – vom Beobachter aus gesehen – hinter den Gasansammlungen (Gaswolken) befindlichen Sternen bemerkbar. Im Radiofrequenzbereich kann dieses Gas sowohl in Absorption als auch in Emission beobachtet werden. Die Ursache für das Nichtleuchten im optischen Bereich ist das Fehlen eines leuchtanregenden Sternes in der Nähe der interstellaren Gaswolke; der Wasserstoff ist deshalb dort auch weitgehend neutral, weshalb man von einem *HI-Gebiet* spricht.

12.3.1.2. Thermodynamischer Zustand des interstellaren Gases. Für den physikalischen Zustand des interstellaren Gases ist das herrschende *interstellare Strahlungsfeld* ausschlaggebend. Es wird von den im interstellaren Raum befindlichen, von den Sternen emittierten Photonen gebildet. Der Beitrag, den ein einzelner Stern leistet, ergibt sich aus seiner Effektivtemperatur, seinem Radius R und seiner

Entfernung r vom betrachteten Ort im interstellaren Raum. Dabei handelt es sich um eine außerordentlich verdünnte Strahlung im Vergleich zu der, die sich in einem Hohlraum konstanter Wandtemperatur (gleich der Effektivtemperatur) befindet, da die strahlende Fläche (Sternscheibe) nur einen verschwindend kleinen Teil der den betrachteten Punkt umgebenden Vollkugel einnimmt. Der *Verdünnungsfaktor* V ergibt sich aus dem Verhältnis der Größe der Sternscheibe zur Oberfläche der Kugel mit dem Radius r, d. h., $V = \pi R^2/(4\pi r^2)$. Im interstellaren Raum ist V im Mittel von der Größenordnung 10^{-16}, doch sind recht große Abweichungen von diesem Wert möglich. Da der Verdünnungsfaktor eine rein geometrische Größe ist, bleibt die Energieverteilung der von den Sternen kommenden Photonen ungeändert, es verkleinert sich allein ihre im betrachteten Punkt verfügbare Anzahl im Vergleich zur Hohlraumstrahlung, also im Vergleich zu den Verhältnissen bei thermodynamischem Gleichgewicht. Zum interstellaren Strahlungsfeld tragen zwar alle Sterne bei, es wird aber vielfach von einem Stern dominiert, und zwar von einem hoher Effektivtemperatur, der dem betrachteten Ort sehr nahe steht. Bei theoretischen Betrachtungen kann man daher i. allg. von diesem Fall ausgehen.

Die Strahlungsverdünnung wirkt sich entscheidend auf den Anregungsgrad der Atome aus. Zu dessen Berechnung kann nicht mehr die Boltzmann-Formel für das thermodynamische Gleichgewicht benutzt werden (Gleichung 12.2.); man muß vielmehr auf ihrer rechten Seite den Verdünnungsfaktor hinzufügen. (Als Temperatur ist die Effektivtemperatur des das Strahlungsfeld am betrachteten Ort bestimmenden Sterns einzusetzen.) Wegen der Kleinheit von V ergibt sich ein entsprechend außerordentlich geringer Anregungsgrad, d. h., sämtliche Atome und Ionen befinden sich praktisch im Grundzustand. Dies Ergebnis läßt sich auch anschaulich machen: Die mittlere Verweilzeit auf einem angeregten Niveau ist um viele Zehnerpotenzen kürzer als die Zeit, die ein Atom im Mittel warten muß, bis es ein zur Anregung geeignetes Photon absorbieren kann. Das Atom geht daher nach jeder Anregung sofort wieder in den Grundzustand über, in dem es lange Zeit auf eine neue Anregung warten muß.

Auch für die Saha-Formel (Gleichung 12.3.) ergeben sich bei verdünnter Strahlung Modifikationen; so muß u.a. auch bei ihr die rechte Seite durch den Verdünnungsfaktor ergänzt werden. Da im interstellaren Raum aber die Anzahldichte der Ionen, entsprechend die der Elektronen, extrem niedrig ist (n_e liegt im Bereich von $\approx 10^{-3}$ bis $10^3\,\mathrm{cm}^{-3}$, ausnahmsweise bei $10^5\,\mathrm{cm}^{-3}$, und steht auf der linken Seite der Saha-Formel), kann der Ionisationsgrad – im Gegensatz zum Anregungsgrad! – durchaus hoch sein. Geringe Elektronendichte bedeutet, daß ein Ion sehr lange warten muß, ehe es rekombinieren kann. Diese Zeit kann wesentlich länger sein als die zwischen Rekombination und folgender Ionisation (obwohl diese auch von der Größenordnung von Jahren ist). Damit existiert das Atom wesentlich im ionisierten Zustand.

Die mittlere kinetische Geschwindigkeit der Gaspartikeln, ausgedrückt durch die *kinetische Temperatur* T_{kin}, wird ebenfalls durch das interstellare Strahlungsfeld bestimmt. Es stellt sich nämlich ein Gleichgewicht zwischen Energiezufuhr infolge von Photoionisation (das befreite Elektron erhält die Differenz zwischen der absorbierten Photonenenergie und der Ionisationsenergie als kinetische Energie) und Energieabfuhr infolge von Umwandlung kinetischer Energie in Anregungsenergie bei unelastischen Stößen ein. Dabei ist die kinetische Temperatur hoch, wenn das Strahlungsfeld genügend Photonen ($\lambda < 91{,}2$ nm) zur Ionisation von Wasserstoff als häufigstes Element hat, hingegen klein, wenn diese Photonen fehlen. In diesem Fall können nämlich nur die (seltenen) Elemente mit einer Ionisationsenergie kleiner als die des Wasserstoffs ionisiert werden. Die Energiezufuhr für das interstellare Gas ist daher entsprechend gering. In jedem Fall bestehen aber starke Unterschiede zwischen T_{eff} des dominierenden Sterns und dem sich einstellenden T_{kin}.

Insgesamt ist somit der physikalische Zustand des interstellaren Gases durch extreme Abweichungen vom thermodynamischen Gleichgewicht charakterisiert, was die theoretische Beschreibung sehr erschwert.

12.3.1.3. Strahlungsprozesse im interstellaren Gas. Im thermodynamischen Gleichgewicht ist die Zahl der Übergänge eines Atoms vom Energiezustand m zum Zustand k genau gleich der Übergänge von k nach m. Bei Abweichungen vom thermodynamischen Gleichgewicht ist das nicht notwendig der Fall. Es muß nur die Gesamtzahl der Übergänge nach k gleich der sein, die von k wegführen, wobei es gleichgültig ist, welche anderen Niveaus und welche Anregungs- bzw. Deaktivierungsmechanismen beteiligt sind. Bei der theoretischen Deutung der interstellaren Spektren ist man daher gezwungen, alle möglichen Prozesse zu untersuchen, die zur Besetzung des Ausgangsniveaus der Linien führen.

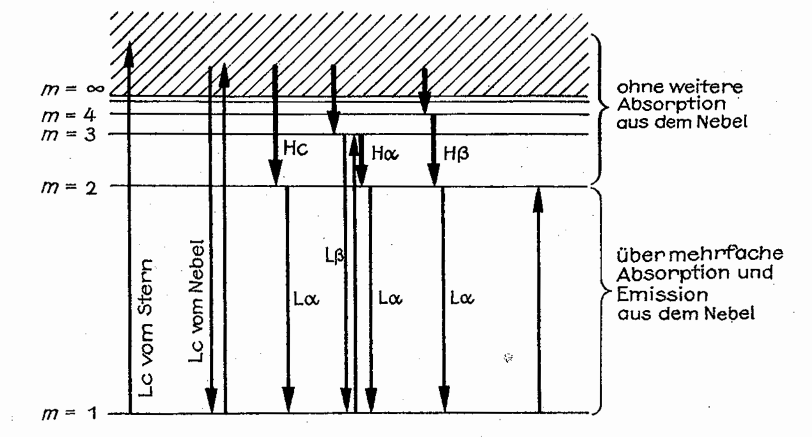


Abb. 12.3.1-1 Schematische Darstellung einiger bei der Rekombination von Elektronen mit Protonen in einem Emissionsnebel auftretenden Übergänge

Emissionsnebel (HII-Gebiete). Ein von einem Stern ausgesandtes Lyman-Kontinuum-Photon (Lc-Quant; $\lambda < 91{,}2$ nm) bewirkt bei der Absorption durch ein Wasserstoffatom dessen Ionisation. Das befreite Elektron verliert einen Teil seiner bei der Ionisation erhaltenen kinetischen Energie durch unelastische Stöße mit anderen Gaspartikeln und bei frei-freien-Übergängen im Felde der vorhandenen Ionen. Die anschließende Rekombination des Elektrons mit einem Proton kann auf das Grundniveau ($m = 1$) erfolgen, wobei wieder ein Lc-Quant emittiert wird, das seinerseits ein anderes Wasserstoffatom ionisieren kann, genau wie das ursprüngliche vom Stern stammende (Abb. 12.3.1–1). Häufiger treten Rekombinationen auf höhere Niveaus auf, wobei jeweils Photonen des zugehörigen Grenzkontinuums (Balmer-Kontinuum (Hc), Paschen-Kontinuum usw.) emittiert werden. Sie können das interstellare Gas praktisch ungehindert durchdringen, da ihre Energie nicht zur Ionisation des Wasserstoffes vom Grundzustand aus ausreicht, und angeregte H-Atome praktisch nicht vorhanden sind. Das Balmer-Kontinuum liefert den Hauptbeitrag zum beobachteten kontinuierlichen Spektrum. Erfolgt die Rekombination auf das erste angeregte Niveau ($m = 2$), so findet anschließend ein Übergang zum Grundzustand mit der Emission eines Lyman-α-Quants (Lα) statt. Dieses kann ein anderes H-Atom aus dem Grundzustand heraus anregen, das dann seinerseits unter Emission von Lα wieder in den Grundzustand zurückkehrt. Im Endeffekt erfolgt also nur eine Streuung der Lα-Strahlung. Bei der Rekombination auf ein noch höheres Niveau kann das Atom kaskadenartig auf $m = 2$ (dabei Emission eines Photons der Balmer-Serie) und danach auf $m = 1$ übergehen (Emission von Lα). Während das Balmer-Quant und alle anderen bei der Kaskade emittierten Photonen vom interstellaren Gas wieder nicht absorbiert werden können, erleidet das Lα-Quant viele Streuprozesse. Insgesamt folgt aus diesen Überlegungen, daß für jedes vom Stern ausgesandte Lc-Photon in einem HII-Gebiet genügender Ausdehnung genau ein Lα- und ein Balmer-Quant (Hα, Hβ, Hγ usw.) entsteht. Aus der Zahl dieser von einem Emissionsgebiet kommenden Photonen kann man daher auf die Zahl der vom leuchtanregenden Stern emittierten Ultraviolettphotonen mit $\lambda < 91{,}2$ nm schließen, wenn man voraussetzt, daß alle diese Photonen auch in diesem Gebiet absorbiert werden. Damit ist eine Möglichkeit zur Bestimmung der Temperatur des leuchtanregenden Sternes gegeben.

Die Rekombination der Elektronen erfolgt auch auf extrem hohe Energieniveaus, wobei m weit über 100 liegen kann. Beim kaskadenartigen Übergang zu tieferen Zuständen wird eine Linienstrahlung großer Wellenlängen emittiert, die u. a. im Radiofrequenzbereich beobachtet werden kann.

Bei genügend hoher Temperatur des leuchtanregenden Sternes ist ein Teil des interstellaren Heliums einmal, ein geringer Teil sogar zweimal ionisiert. Bei der Rekombination der Elektronen mit den Heliumionen kommt es zu einer entsprechenden kontinuierlichen und Linienstrahlung wie bei der Rekombination mit den Protonen. Das gleiche gilt für die anderen vorhandenen Elemente, nur sind wegen deren geringer Häufigkeit die Linien in den beobachteten Spektren entsprechend schwach.

In den Spektren von Emissionsnebeln fallen Linien auf, die ihre Stärke weniger den herrschenden physikalischen Bedingungen als mehr der zufälligen Koinzidenz bestimmter Spektrallinien verdanken, die von unterschiedlichen Atomen bzw. Ionen ausgehen. Beispielsweise entsteht als Folge der Rekombination von Elektronen mit He^{++}-Ionen eine

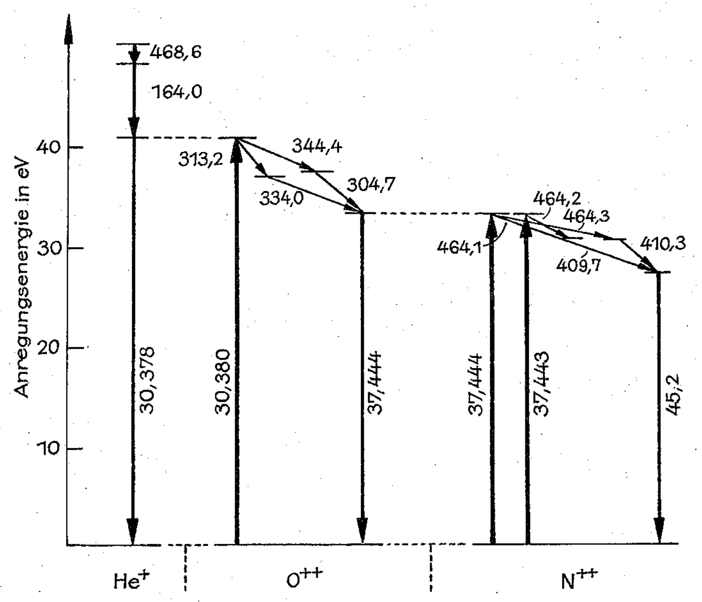


Abb. 12.3.1–2 Ausschnitte aus den Termschemata der Ionen He^{+}, O^{++} und N^{++}. Angegeben sind einige Strahlungsübergänge mit den zugehörigen Wellenlängen (in Nanometer). Die vom He^{+}-Ion emittierte Strahlung mit $\lambda = 30{,}378$ nm kann vom O^{++}-Ion absorbiert werden, das zufälligerweise ein Resonanzniveau mit der (fast) genau richtigen Anregungsenergie besitzt. Die notwendigerweise bei $\lambda = 37{,}444$ nm auftretenden Emissionen des so angeregten O^{++}-Ions können nun ihrerseits N^{++}-Ionen anregen

Spektrallinie bei $\lambda = 30{,}378$ nm, die mit einer Resonanzlinie von O^{++} koinzidiert (Resonanzlinien gehen vom Grundzustand aus), also von diesem Ion absorbiert werden kann (Abb. 12.3.1–2). Bei dessen Übergang in den Grundzustand wird u. a. eine Linie bei $\lambda = 33{,}44$ nm emittiert, die zur Anregung bestimmter Energieniveaus von N^{++} geeignet ist. Geht dieses Ion schließlich in den Grundzustand zurück, treten Linien im visuellen und ultravioletten Spektralbereich auf, die bei alleiniger Anwesenheit von Stickstoff unter thermodynamischen Gleichgewichtsbedingungen ganz schwach wären.

Die in den HII-Gebieten vorhandenen Elektronen haben genug kinetische Energie, um tiefliegende Energieniveaus in Atomen und Ionen durch Stoß anzuregen. Dabei handelt es sich i. allg. um metastabile Niveaus, für die Strahlungsübergänge zu tieferliegenden Zuständen nach den üblichen Auswahlregeln verboten sind. Die Verweilzeiten sind damit entsprechend lang; sie können Sekunden bis Monate betragen (Abb. 12.3.1–3), im Gegensatz zu den $\approx 10^{-8}$ s bei normalen Niveaus. Unter Laborbedingungen treten diese *verbotenen Linien* nicht auf, da während der großen Verweilzeiten ein angeregtes Atom oder Ion sehr viele Stöße erleidet, so daß eine weitere Stoßanregung oder -deaktivierung, aber praktisch kein Strahlungsübergang erfolgt. In den HII-Gebieten ist hingegen die Teilchendichte so niedrig, daß die mittlere Zeit zwischen zwei Stößen eines Partikels wesentlich länger als die Verweilzeit auf dem metastabilen Niveau ist. Jeder Stoßanregung folgt daher praktisch immer ein Strahlungsübergang. Die verbotenen Linien zählen trotz der geringen relativen Häufigkeit der Elemente, von denen sie ausgehen, zu den stärksten der Emissionsnebel.

Die Anregung metastabiler Niveaus infolge von Elektronenstoß führt dazu, daß die kinetische Temperatur der Elektronen (*Elektronentemperatur* T_e) in den HII-Gebieten wesentlich unter der Effektivtemperatur des leuchtanregenden Sterns liegt, daß sie aber innerhalb eines Emissionsgebietes ziemlich konstant ist. Sinkt T_e zu stark ab, wird entsprechend die mittlere Elektronengeschwindigkeit, damit die Zahl der Stöße, verringert und demzufolge auch die je Zeit- und Volumeneinheit durch Stoß abgeführte Energie; die Energiezuführung über Strahlungsionisation bleibt hingegen ungeändert. Die Elektronentemperatur der HII-Gebiete liegt im Bereich von $\approx 7\,000$ bis $10\,000$ K.

HI-Gebiete. Im optischen Spektralbereich macht sich das HI-Gas durch Absorptionslinien bemerkbar, die dem Spektrum der – vom Beobachter aus gesehen – hinter den Wolken neutralen Wasserstoffs gelegenen Sternen zusätzlich zu den stellaren Linien aufgeprägt sind. Die interstellaren Linien haben wegen der geringen kinetischen Temperatur des HI-Gases und der daher geringen Doppler-Verbreiterung (vgl. 12.2.2.3.) eine sehr geringe Breite verglichen mit den stellaren Linien. Es können aber nur solche Atome und Ionen beobachtet werden, die Resonanzlinien im optischen Spektralbereich haben (z. B. Na, Ca, Ca^+, Fe). Wasserstoff, obwohl häufigstes Element, gehört nicht dazu. Seine Resonanzlinien (Lyman-Serie) liegen, wie die der meisten anderen Atome und Ionen, im Ultraviolettbereich und können daher nur mittels extraterrestrischer Beobachtungen untersucht werden.

Neutraler interstellarer Wasserstoff wird hauptsächlich im Radiofrequenzbereich beobachtet. Der Grundzustand des Wasserstoffatoms hat nämlich eine Hyperfeinstruktur, die durch die zwei Einstellmöglichkeiten des Elektronenspins zum Kernspin (parallel oder antiparallel) bedingt ist. Beim

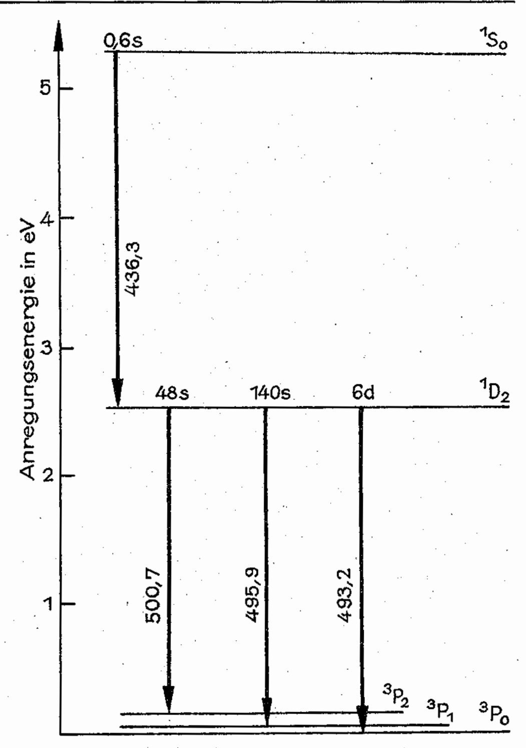


Abb. 12.3.1–3 Die Aufspaltung des Grundzustandes des O^{++}-Ions in metastabile Energieniveaus (Termbezeichnung rechts; die Energiedifferenzen der ^3P-Terme sind um den Faktor 4 überhöht dargestellt). Für die einzelnen Übergänge ist neben der Wellenlänge (in Nanometer) auch die entsprechende mittlere Verweilzeit im angeregten Niveau angegeben

Übergang vom parallelen zum antiparallelen Zustand wird eine Emissionslinie bei $\lambda = 21{,}1$ cm ausgestrahlt, beim umgekehrten Übergang absorbiert. Dabei handelt es sich um einen extrem verbotenen Übergang. Die mittlere Verweilzeit im angeregten Zustand beträgt $\approx 11 \cdot 10^6$ Jahre. Wegen der geringen Teilchendichte (Mittelwert 1 bis 10 H-Atome je cm³) und der geringen kinetischen Temperatur im HI-Gas (im Mittel 80 bis 100 K) ist die Zeit zwischen zwei Stößen eines H-Atoms zwar viel kürzer als die Verweilzeit, aber doch noch so groß, daß eine merkliche Anzahl von Strahlungsübergängen erfolgen kann. Diese *21-cm-Linie* wird in Absorption beobachtet, wenn sich das HI-Gas zwischen dem Beobachter und einer Radiokontinuumstrahlungsquelle befindet. Aus der Doppler-Breite der Linie (vgl. 12.2.2.3.) kann auf die kinetische Temperatur des Wasserstoffgases, aus der Gesamtabsorption bzw. -emission auf die Zahl der in der Sichtlinie befindlichen Wasserstoffatome, damit auf ihre räumliche Dichte, geschlossen werden.

Molekülwolken. In sehr dichten HI-Gebieten können sich *interstellare Moleküle* bilden, von denen gegenwärtig ≈ 50 bekannt sind, wobei es sich um Verbindungen der Elemente Wasserstoff, Kohlenstoff, Stickstoff, Sauerstoff, Silicium und Schwefel handelt. Einige wenige (CH, CH^+, CN) werden im optischen Spektralbereich, das H_2-Molekül im

Ultraviolettbereich mittels Absorptionslinien, die überwiegende Mehrzahl aber mittels Emissionslinien im Radiofrequenzbereich beobachtet (in Absorption nur, wenn die Molekülwolke auf eine Radiokontinuumquelle projiziert erscheint). Der wesentliche Strahlungsmechanismus bei den Emissionslinien ist der Übergang von einem höheren Rotationsniveau zu einem niedrigeren (z. B. bei CO, HCN, HC_3N). Die Anregung auf das höhere Niveau erfolgt bei Stößen vor allem durch H_2-Moleküle. Bei bestimmten Molekülen können Rotationslinien noch eine zusätzliche Linienaufspaltung infolge der innermolekularen Rotation bestimmter Atomgruppen (z. B. der CH_3-Gruppe in CH_3OH oder CH_3CN) zeigen. Darüber hinaus gibt es noch weitere Linienaufspaltungen, die z. B. auf die Wechselwirkung von Molekülrotation und Elektronen- oder Kernspin zurückzuführen sind. Die Identifikation der beobachteten Linien bereitet zum Teil erhebliche Schwierigkeiten, da der Vergleich mit Laborspektren schwer oder unmöglich ist (z. B. bei Radikalen), und da auch die theoretische Berechnung der Spektren von Vielteilchenmolekülen kompliziert ist.
Aus der Doppler-Breite der Moleküllinien läßt sich die kinetische Temperatur herleiten; sie liegt im Mittel bei 30 bis 50 K, doch werden auch Werte von 5 bis 10 K gemessen. Die Teilchendichte erhält man aus den Linienstärken. Für die H_2-Moleküle ergeben sich Werte zwischen 10^4 bis $10^7\ cm^{-3}$; die der anderen Moleküle ist um viele Größenordnungen geringer. Auf $\approx 10^4$ H_2-Moleküle kommt im Durchschnitt z. B. nur ein CO-Molekül (das nächst häufigste nach H_2) und nur 0,01 OH- oder NH_3-Moleküle.

gleich zum thermodynamischen Gleichgewichtszustand verbunden.
Damit die interstellaren Moleküle existieren können und nicht durch das interstellare Strahlungsfeld dissoziiert werden, ist es notwendig, daß dieses genügend stark abgeschirmt wird. Dafür sorgt der interstellare Staub, dessen Teilchendichte proportional der der Gaspartikeln ist und daher in den Molekülwolken sehr hohe Werte erreicht. Die Staubteilchen haben gleichzeitig für die Molekülentstehung eine große Bedeutung. Die Bildung einfacher (zweiatomiger) Moleküle direkt aus der Gasphase ist nämlich wegen der geringen Gasdichte nicht möglich, da die Bindungsenergie der Moleküle nicht schnell genug abgeführt werden kann, diese also sofort wieder zerfallen. Erfolgt die Molekülbildung hingegen an der Oberfläche von Staubpartikeln, so kann die Bindungsenergie auf diese übergehen und gegebenenfalls gleich zur Ablösung des neu entstandenen Moleküls von der Oberfläche genutzt werden. Anderenfalls sorgt die in den dichten Wolken noch vorhandene Reststrahlung des interstellaren Strahlungsfeldes für eine Ablösung der Moleküle von den Staubteilchen. Mehratomige Moleküle dürften hingegen im wesentlichen in der Gasphase beim direkten Stoß von wenigatomigen Molekülen mit Ionen oder ionisierten Radikalen gebildet werden, da in diesem Fall viele Energiezustände zur Verfügung stehen, auf die die freiwerdende Bindungsenergie verteilt werden kann. Beobachtet werden z. B. so vielatomige Moleküle wie CH_3CH_2OH (Ethylalkohol) oder HC_9N (Cyanooktatetrain).

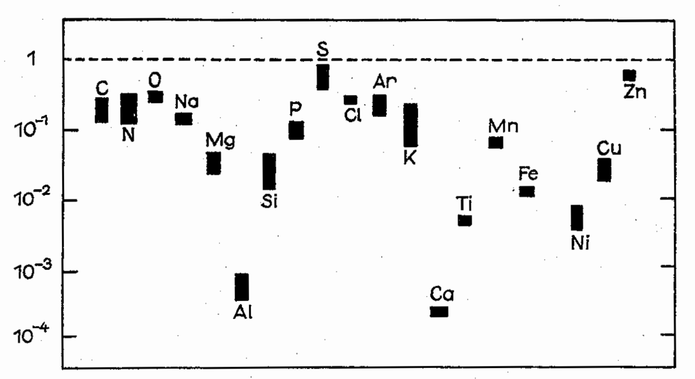


Abb. 12.3.1–4 Relative Elementenhäufigkeit im interstellaren Gas in Richtung auf den Stern ζ Ophiuchi im Vergleich zu den Sonnenwerten. Die Balkenlänge gibt die Unsicherheit der Bestimmung an. Werte unterhalb der gestrichelten Linie bedeuten, daß das betreffende Element im Vergleich zur Sonne unterhäufig ist

Die Stärke der Emissionslinien wird durch die kinetische Temperatur, die Dichte der Gaspartikeln sowie durch die Größe des Stoßanregungskoeffizienten der Moleküle bestimmt. Sind diese Größen bekannt, so kann man das Intensitätsverhältnis unterschiedlicher Linien, die vom gleichen Molekül emittiert werden, berechnen. Bei einigen Molekülen (OH, H_2O, SiO) treten zum Teil erhebliche Abweichungen zwischen den beobachteten und den theoretisch erwarteten Intensitätsverhältnissen auf. Dies ist auf die Wirkung eines *interstellaren Maserprozesses* zurückzuführen, wobei die Pumpenenergie aus der Strahlungsenergie benachbarter Sterne oder Infrarotquellen stammt. Der Masermechanismus ist mit einer extremen Abweichung der Besetzungsdichte der beteiligten Energieniveaus im Ver-

12.3.1.4. Chemische Zusammensetzung des interstellaren Gases. Zur *quantitativen Spektralanalyse des interstellaren Gases* benutzt man im Prinzip die gleichen Methoden wie bei den Sternatmosphären (vgl. 12.2.2.4.). Aus den beobachteten Spektren werden für möglichst viele Linien eines Elements Äquivalentbreiten abgeleitet, aus denen dann mit Hilfe von Wachstumskurven die Zahl der in der Sichtlinie liegenden Atome bestimmt wird.
Ganz allgemein findet man, daß die chemische Zusammensetzung des interstellaren Gases etwa der mittleren kosmischen Elementenhäufigkeit entspricht (vgl. 12.2.2.4.). In verschiedenen HI-Gebieten findet man aber auch, daß die Häufigkeit einzelner schwerer Elemente bis zu einem Faktor 1000 gegenüber dem Sonnenwert reduziert ist,

Abb. 12.3.1–4. Dies ist wahrscheinlich darauf zurückzuführen, daß diese Elemente bevorzugt an interstellare Staubteilchen anlagern und so dem interstellaren Gas entzogen werden (vgl. 12.3.2.2.).

12.3.2. Interstellarer Staub

12.3.2.1. Beobachtungen. Der interstellare Staub kann sich im optischen Spektralbereich dadurch bemerkbar machen, daß er das Licht benachbarter Sterne geringer Effektivtemperatur ($T_{eff} \lesssim 20000$ K) streut. Man beobachtet dann sogenannte *Reflexionsnebel*. Das Spektrum des gestreuten Lichtes unterscheidet sich praktisch nicht vom Sternspektrum, man beobachtet also ein Kontinuum mit überlagerten Absorptionslinien. Im Infrarotbereich kann sich interstellarer Staub durch eine thermische Kontinuumsstrahlung bemerkbar machen. Es handelt sich dabei um die Wärmestrahlung, die von Staubteilchen ausgesandt wird, die durch Absorption von Sternlicht auf einige 100 K aufgeheizt wurden. Gelegentlich sind dem Infrarotkontinuum einzelne breite Emissionsbanden überlagert, z. B. bei 9,7 µm und bei 18 µm.
Nichtleuchtender interstellarer Staub verursacht eine allgemeine Schwächung des Lichtes derjenigen Sterne, die – vom Beobachter aus gesehen – hinter einer Staubwolke stehen. Diese *interstellare Extinktion* ist wellenlängenabhängig, kurzwelliges Licht wird stärker geschwächt als längerwelliges (Abb. 12.3.2–1). Im optischen Spektralbereich gilt näherungsweise $A(\lambda) \sim \lambda^{-1}$, wenn $A(\lambda)$ den Extinktionsbetrag bei der Wellenlänge λ bezeichnet. Der Effekt bewirkt somit eine *interstellare Verfärbung* (Verrötung) weitentfernter Sterne. Zu dieser kontinuierlichen Extinktion können noch einzelne Absorptionsbanden hinzukommen; beobachtet wird z. B. eine im Ultraviolettbereich bei 220 nm ($\lambda^{-1} = 4{,}5\ \mu m^{-1}$) oder im Infrarotbereich bei 3,07 µm und (entsprechend den Emissionsbanden) bei 9,7 µm und 18 µm.
12.3.2.2. Zustand der interstellaren Staubteilchen. Die Theorie des interstellaren Staubes versucht, aus den von Staubteilchen verursachten Beobachtungseffekten auf die Größe und chemische Beschaffenheit der Teilchen zu schließen. Dies ist erheblich schwieriger als im Fall des interstellaren Gases, da die optischen Eigenschaften von Festkörperpartikeln wesentlich komplizierter sind als die von Atomen oder Molekülen. Hinweise auf die chemisch-mineralogische Beschaffenheit des interstellaren Staubes gewinnt man aus den in den verschiedenen Spektralbereichen auftretenden Emissions- und Absorptionsbanden. Dabei sind für die mineralogische Identifikation die Lage des Bandenmaximums und das Bandenprofil von Interesse. Erschwerend für das Verständnis ist die Strukturlosigkeit der beobachteten Spektralmerkmale, was darauf hindeutet, daß die Staubteilchen keine geordnete kristalline Struktur aufweisen, sondern amorph aufgebaut sein dürften. Die Banden bei 9,7 und 18 µm werden wahrscheinlich durch Silikatteilchen, die bei 3,07 µm durch Wassereis verursacht. Die Entstehung der Ultraviolettabsorption bei 220 nm ist noch umstritten, sie könnte von kleinen Graphitteilchen herrühren.
Hinweise auf den mittleren Teilchenradius erhält man z. B. aus der Wellenlängenabhängigkeit der kontinuierlichen Extinktion im optischen Spektralbereich. Danach muß der Teilchenradius von der Größenordnung der Wellenlänge des extingierten Lichtes sein, da sehr kleine Teilchen eine λ^{-4}-Abhängigkeit, sehr große eine wellenlängenunabhängige Extinktion hervorrufen. Die Radienverteilung (≈10 bis einige 100 nm) läßt sich aber nicht direkt aus den Beobachtungen ableiten. Man muß vielmehr von theoretischen Ansätzen ausgehen und sehen, ob sich diese mit den Beobachtungen vertragen.

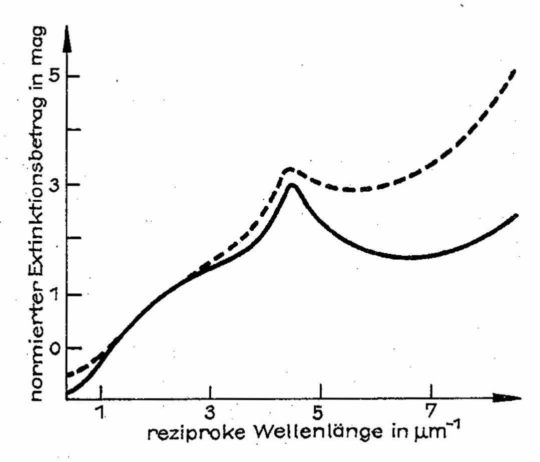


Abb. 12.3.2–1 Die interstellare Extinktion in Richtung auf verschiedene Sterne (– – – ζ Ophiuchi, —— τ Scorpii) in Abhängigkeit von der Wellenlänge, wobei über der reziproken Wellenlänge der (normierte) Extinktionsbetrag aufgezeichnet wurde. Die Stärke der Extinktion bei 220 nm ($\lambda^{-1} = 4{,}5\ \mu m^{-1}$) variiert von Stern zu Stern

Es zeigt sich, daß sich die Beobachtungen in ihrer Gesamtheit nicht durch ein einziges Staubteilchenmodell erklären lassen, man muß mehrere, chemisch unterschiedliche Teilchensorten mit jeweils unterschiedlichen Radienverteilungen annehmen, wobei die einzelnen Modellparameter so gewählt werden, daß die Beobachtungen möglichst genau reproduziert werden. Dabei ergibt sich, daß man dies mit recht unterschiedlichen Parametersätzen erreichen kann, so daß gegenwärtig keine eindeutigen Aussagen über die Beschaffenheit des interstellaren Staubes möglich sind.
Da man die chemische Zusammensetzung der Staubteilchen praktisch nicht kennt, läßt sich auch nichts Konkretes über ihre Entstehung sagen. Es wird angenommen, daß sich die Teilchen u. a. durch Kondensation von Gas in den äußeren Atmosphärenschichten kühler Riesensterne und in den bei Supernovaausbrüchen abgestoßenen Gasmassen bilden. Durch die in diesem Gas vorhandene Häufigkeit der verschiedenen Elemente sowie deren Kondensationsverhalten unter den herrschenden physikalischen Bedingungen kann man näherungsweise auf die chemische Struktur der entstehenden Partikeln schließen. Danach dürften vor allem Silikat- und Graphitteilchen vorhanden sein. Im interstellaren Raum können sich weitere Gaspartikeln an die vorhandenen Staubteilchen anlagern, was ein zusätzliches Teilchenwachstum bedeutet und zur chemischen Inhomogenität führt; die hinzukommenden Schichten dürften nämlich hauptsächlich aus leichtkondensierbaren Materialien, wie Wassereis, bestehen. Eine teilweise Verkleinerung tritt dann ein, wenn solche Staubteilchen einem starken interstellaren Strahlungsfeld ausgesetzt werden, da dann einzelne Atome und Moleküle von der Teilchenoberfläche verdampfen können.

Das Massenverhältnis von interstellarem Staub zu interstellarem Gas liegt im Mittel bei $\approx 1:100$, wobei aber gebietsweise große Abweichungen davon beobachtet werden. Größenordnungsmäßig kommt auf $\approx 10^{10}$ bis 10^{12} interstellare H-Atome ein Staubpartikel.

12.3.3. Kosmische Strahlung

Die kosmische Strahlung ist eine aus Atomkernen und Elektronen bestehende hochenergetische Teilchenstrahlung im interstellaren Raum. Die kinetische Energie der Atomkerne reicht von $\approx 10^{-12}$ bis 10 J je Teilchen, doch nimmt ihre Häufigkeit mit steigender Energie stark ab. (Bei einer Temperatur von 10^9 K beträgt die mittlere kinetische Energie der Gaspartikeln nur $2 \cdot 10^{-13}$ J!).
Unter den Atomkernen dominieren mit etwas mehr als 90 % die Protonen gefolgt von Heliumkernen; alle schwereren machen weniger als 1 % aus. Die Elemente Lithium, Beryllium und Bor haben eine $\approx 10^6$-mal größere relative Häufigkeit als in der Sonnenatmosphäre. Das ist darauf zurückzuführen, daß diese Kerne erst im interstellaren Raum durch Zertrümmerung von schweren Atomkernen, vor allem Eisen, durch hochenergetische Protonen und Heliumkerne entstehen. Am eigentlichen Ursprungsort der kosmischen Strahlung dürften diese Elemente gleiche (sehr geringe) relative Häufigkeit haben wie sonst beobachtet. Die niederenergetischen Partikeln der kosmischen Strahlung verursachen beim Zusammentreffen mit interstellaren Atomen deren Ionisation. Da sie auch dichte Gaswolken durchdringen, ist ein Teil des Wasserstoffes (wenige Prozent) auch in den HI-Gebieten ionisiert, wozu das normale interstellare Strahlungsfeld nicht in der Lage ist (vgl. 12.3.1.2.). Die kosmische Strahlung trägt damit zur Aufheizung des interstellaren Gases bei.
Der Ursprung der kosmischen Strahlung ist nicht sicher bekannt. Ein Teil entsteht wahrscheinlich bei Supernovaausbrüchen. Eine andere Quelle scheint das Zentralgebiet des Milchstraßensystems zu sein. Bei den Teilchen höchster Energie handelt es sich vielleicht um Partikeln, die aus den aktiven Kerngebieten anderer Sternsysteme stammen. Die eigentliche Beschleunigung, durch die die Atomkerne ihre extrem hohe kinetische Energie erlangen, findet aller Wahrscheinlichkeit nach in starken, inhomogenen Magnetfeldern statt. Die kosmische Strahlung macht sich beobachtungsmäßig von der Erde aus vor allem dadurch bemerkbar, daß infolge von Wechselwirkungen mit Atomkernen der Erdatmosphäre sich eine Fülle komplizierter Umwandlungsprozesse ergibt. Dabei kann ein einzelnes, primäres Teilchen einen ganzen Schauer von Sekundärteilchen und hochenergetischen Photonen auslösen, wobei die Intensität dieser Sekundärstrahlung proportional der Energie des auslösenden Primärteilchens ist.

12.4. Physikalischer Zustand der Welt als Ganzes

Die überwiegende Masse der beobachteten Materie existiert in Form von Sternen. Diese bilden Sternsysteme (eines ist das Milchstraßensystem, dem die Sonne samt dem Planetensystem angehört), die zwischen einigen 10^8 bis $\approx 10^{12}$ Einzelsterne enthalten. Die interstellare Materie trägt – grob gemittelt – nur ≈ 5 bis 10 % zur Gesamtmasse der Systeme (Galaxien) bei. Die Galaxienhaufen (Mitgliederzahl zwischen ≈ 50 und mehreren 1000) können als die größten beobachteten Strukturelemente im Weltall aufgefaßt werden.
Die *Kosmologie* als Teilgebiet der Astronomie untersucht die Struktur – auch in ihrer zeitlichen Veränderung – der Welt als Ganzes. Von dieser ist der Beobachtung prinzipiell nur ein räumlicher Teilbereich zugänglich, von dem man i. allg. nicht weiß, welchen Bruchteil des Weltganzen er ausmacht. Auch zeitliche Strukturänderungen sind – wenn überhaupt – nur für eine endliche Zeitspanne in der Vergangenheit beobachtungsmäßig verfolgbar. Alle auf Beobachtungen beruhenden Aussagen, die die Welt als Ganzes betreffen, stellen daher Verallgemeinerungen dar, wobei angenommen wird, daß der beobachtete Teilbereich typisch für das Weltganze ist. Ohne diese Annahme wäre jede kosmologische Theorie wertlos, da sie nicht beobachtungsmäßig gestützt und überprüfbar wäre.
Die wissenschaftliche Grundlage der Kosmologie bilden neben den astronomischen Beobachtungen die bekannten Gesetze der Physik. Es lassen sich daher wissenschaftlich begründet nur solche Materiezustände im Weltall beschreiben, für die die physikalischen Gesetze bekannt sind, denn nur für sie ist die Beweis- und Überprüfbarkeit gesichert.

12.4.1. Beobachtungsgrundlagen

Bei der Untersuchung der großräumigen Struktur des Kosmos sieht man von Detailstrukturen ab. Sie werden als lokale Störungen betrachtet. Dies betrifft insbesondere die Materieverteilung. Vereinfachend stellt man sich die Sternsysteme als Testpartikeln vor, die gegenüber ihrer lokalen Umgebung ruhen, da ihre Individualgeschwindigkeiten gering sind.
12.4.1.1. Hubble-Effekt. In den Spektren weit entfernter Sternsysteme sind die Absorptionslinien gegenüber den im Labor bestimmten Wellenlängen nach größeren Wellenlängen verschoben, wobei die Verschiebung mit der Entfernung r der Systeme zunimmt *(Rotverschiebung)*. Dies bedeutet eine radiale Bewegung vom Beobachter weg. Für die Radialgeschwindigkeit v der Systeme ergibt sich mithin $v = H_0 r$ *(Hubble-Beziehung)*. Der Proportionalitätsfaktor H_0 *(Hubble-Parameter)* ist dabei unabhängig von der Beobachtungsrichtung. Sein Wert ist wegen der Schwierigkeiten bei der Entfernungsbestimmung von Galaxien nicht sehr sicher bekannt; er liegt zwischen 50 und 100 $\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$, wobei vielfach mit $H_0 = 55\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ gerechnet wird ($1\,\text{Mpc} = 3{,}0856 \cdot 10^{22}\,\text{m}$). Im überblickbaren Teil des Weltalls beobachtet man also eine systematische und isotrope Expansion des Raumes, die eine Vergrößerung aller Abstände zur Folge hat. Die Lage des Milchstraßensystems im Zentrum der Expansion ist nur scheinbar, da bei solch einer Expansion jeder Beobachter das gleiche Erscheinungsbild wahrnimmt; es ist kein Punkt und keine Richtung ausgezeichnet.
Bei einer isotropen Expansion des Raumes sind alle Abstände r in gleicher Weise variabel, $r = x_0 R(t)$. Dabei ist x_0 eine dimensionslose, den Abstand zweier Raumpunkte zu einem bestimmten Zeitpunkt t_0 charakterisierende konstante Größe und $R(t)$ eine von der Zeit t abhängige Funktion mit der Dimension einer Länge. Bezeichnet t_0 die Gegenwart, so ergibt sich $H_0 = \dot{R}(t_0)/R(t_0)$, wobei $\dot{R}(t_0)$ die gegenwärtige Expansionsgeschwindigkeit bedeutet. Der Hubble-Parameter hat die Dimension einer reziproken Zeit. Mit $H_0 = 55\,\text{km}\,\text{s}^{-1}\,\text{Mpc}^{-1}$ erhält man $H_0^{-1} = T_H \approx 18 \cdot 10^9$ Jahre. Es ist dies die Zeit *(Hubble-Zeit)*, die

bei der gegenwärtigen, als konstant angenommenen Expansionsgeschwindigkeit notwendig war, damit die Sternsysteme ihre jetzigen gegenseitigen Abstände erreichen konnten. Vor T_H Jahren waren alle Abstände extrem klein. Die gesamte Materie war auf ein verschwindend kleines Volumen konzentriert, so daß die Dichte jede Grenze überschritt. Dieser Zeitpunkt ist somit durch einen singulären Zustand *(kosmische Singularität)* gekennzeichnet.

12.4.1.2. Materiedichte. Die mittlere (gegenwärtige) Materiedichte ϱ_0 im Weltall ergibt sich aus der in einem bestimmten Volumen befindlichen Anzahl von Sternsystemen und der mittleren Masse der Systeme. Um einen repräsentativen Wert für ϱ_0 zu erhalten, müssen die Abmessungen der betrachteten Volumina größer als die charakteristischen Lineardimensionen der größten Struktureinheiten im Kosmos, also der Galaxienhaufen, sein. Man muß mithin sehr weit entfernte Sternsysteme in die Beobachtungen einbeziehen. Ihre Entfernung läßt sich aber nur über die Hubble-Beziehung ermitteln, so daß der abgeleitete Wert von ϱ_0 auch von H_0 abhängt. Nimmt man wieder $H_0 = 55\,\mathrm{kms^{-1}\,Mpc^{-1}}$ an, so erhält man für die mittlere Materiedichte im Weltall $3 \cdot 10^{-28}$ bis $6 \cdot 10^{-28}\,\mathrm{kg/m^3}$, was $\approx 0{,}2$ bis 0,4 H-Atome/m³ entspricht. Diese Werte stellen untere Grenzwerte dar, da in großen Entfernungen nur leuchtkraftstarke Sternsysteme beobachtbar sind. Es kann sich ein nicht verschwindend kleiner Bruchteil der Materie z. B. in Form von leuchtkraftschwachen (damit massearmen) Zwergsystemen der Beobachtung entziehen.

Da alle Lineardimensionen mit $R(t)$ variieren, gilt für die Materiedichte $\varrho(t) \sim \varrho_0/R^3(t)$, wobei ϱ_0 die gegenwärtige mittlere Dichte bedeutet.

12.4.1.3. Kosmische Hintergrundstrahlung. Das Weltall ist gleichmäßig mit einer Strahlung erfüllt, deren spektrale Energieverteilung der eines Hohlraumstrahlers der Temperatur 2,79 K gleicht (*3-K-Strahlung*, Abb. 12.4.1–1). Da es sich um eine rein kontinuierliche Strahlung von außerordentlicher Isotropie handelt, kann sie von keiner Kategorie bekannter Objekte (Sterne, interstellare Materie) ausgehen; es muß sich um eine aus den Tiefen des Kosmos stammende Raumstrahlung *(kosmische Hintergrundstrahlung)* handeln. Für die Energiedichte einer Hohlraumstrahlung gilt $u_s = 4\sigma T^4/c$; mit $T = 2{,}79\,\mathrm{K}$ ergibt sich für die Gegenwart $u_{s0} = 4{,}6 \cdot 10^{-14}\,\mathrm{J/m^3}$. Wegen der Energie-Masse-Äquivalenz läßt sich diese Energiedichte in eine äquivalente Massendichte umrechnen, $\varrho_{s0} = u_{s0}/c^2 = 5 \cdot 10^{-31}\,\mathrm{kg/m^3}$. Wegen ihrer Kleinheit spielt sie – gegenwärtig – gegenüber der mittleren Materiedichte keine Rolle.

Auch die Strahlungsdichte ist zeitlich variabel und zwar gilt $\varrho_s \sim \varrho_{s0}/R^4(t)$, da einerseits die Zahl der Protonen je Volumeneinheit wie $1/R^3(t)$ variiert und andererseits jedes einzelne Photon wegen der Raumexpansion (Rotverschiebung) noch eine Energieabnahme wie $1/R(t)$ erleidet. Aus der Formel für die Energiedichte einer Hohlraumstrahlung ergibt sich daher für die Temperatur T_s der kosmischen Hintergrundstrahlung $T_s(t) \sim 1/R(t)$.

12.4.2. Weltmodelle

12.4.2.1. Weltpostulate. Zur wissenschaftlichen Beschreibung des Gesamtweltalls konstruiert man Weltmodelle, die in sich widerspruchsfrei und so beschaffen sein müssen, daß eine umkehrbar eindeutige Abbildung des beobachteten Teilbereiches des Weltalls auf den entsprechenden Teilbereich des Modells möglich ist. Im gegenwärtigen Zustand

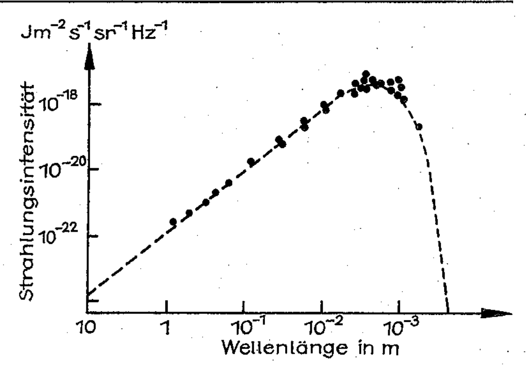


Abb. 12.4.1–1 Die beobachtete Intensitätsverteilung der kosmischen Hintergrundstrahlung (Punkte) und das Spektrum eines Hohlraumstrahlers der Temperatur 2,8 K (ausgezogene Linie)

des Kosmos spielt für seine Struktur im Großen von allen möglichen Wechselbeziehungen zwischen seinen Bestandteilen die Gravitation die entscheidende Rolle. Da die allgemeine Relativitätstheorie die beste bekannte Gravitationstheorie ist, wird sie den meisten kosmologischen Theorien zugrundegelegt. In ihr besteht zwischen Gravitation (hervorgerufen durch die Massenverteilung) und geometrischer Struktur des Raumes (seiner Metrik) ein untrennbarer Zusammenhang, der durch die sogenannten Einsteinschen Feldgleichungen beschrieben wird. Bei ihrer Anwendung auf die konkrete Weltallstruktur bedarf es außer Anfangsbedingungen noch zusätzlicher Bedingungen, die in Form von Postulaten formuliert werden. Sie stellen Extrapolationen von Beobachtungsergebnissen auf den Gesamtkosmos dar.

Das *Homogenitätspostulat* fordert, daß jeder relativ zu seiner Umgebung ruhende Beobachter vom Weltall in allen seinen Eigenschaften den gleichen Anblick hat wie jeder beliebige andere, zum gleichen Zeitpunkt auch gegenüber seiner Umgebung ruhende Beobachter. Das *Isotropiepostulat* fordert, daß der Anblick der Welt im Großen in jeder Richtung gleich sein soll. Diese Postulate beinhalten u. a. die universelle Gleichartigkeit der Materie samt der durch sie gebildeten Strukturen (Sterne, Sternsysteme usw.), die Gleichartigkeit der Bewegungsverhältnisse sowie die universelle Gültigkeit der Naturgesetze.

12.4.2.2. Raumstruktur. Mit der Einführung der Weltpostulate in die Theorie wird eine (großräumig) gleichmäßige Materieverteilung angenommen, was eine entsprechend einfache Raumstruktur der Weltmodelle bedingt. Dabei braucht der dreidimensionale Raum nicht notwendig die von der Anschauung her gewohnte euklidische Struktur aufzuweisen. Er kann gekrümmt sein, nur muß die Krümmung zu einem bestimmten Zeitpunkt überall gleich sein. Prinzipiell sind sowohl positiv gekrümmte Räume (sphärische) als auch negativ gekrümmte (hyperbolische) und flache (euklidische) Räume möglich. Abb. 12.4.2–1 gibt Beispiele von gekrümmten zweidimensionalen Räumen. Die mathematische Analyse zeigt, daß man die Funktion $R(t)$ in diesem Zusammenhang als Krümmungsradius des Raumes interpretieren kann, wobei der zeitliche Verlauf von $R(t)$ sich als Lösung der – aufgrund der Weltpostulate stark vereinfachten – Einsteinschen Feldgleichungen ergibt. Unterschiedliche Weltmodelle sind durch unterschied-

liche Verläufe $R(t)$ charakterisiert. Aus der Vielzahl der Möglichkeiten wird das tatsächlich im Weltall realisierte Modell durch die als Modellparameter fungierenden (bzw. als Anfangswerte in die Gleichungen eingehenden) Beobachtungsgrößen ϱ_0 und H_0 ausgewählt. Ihre durch die Beobachtungsmöglichkeiten verursachten Ungenauigkeiten bedingen ihrerseits Ungenauigkeiten, mit denen die berechneten Weltmodelle behaftet sind.

so war $R(t)$ früher kleiner, mithin sowohl die Materie- als auch die Strahlungsdichte (bzw. ihre äquivalente Massendichte) größer. Dabei wächst wegen der starken $R(t)$-Abhängigkeit ϱ_s schneller als ϱ; für einen Zeitpunkt t_s galt $\varrho(t_s) = \varrho_s(t_s)$, während noch früher $\varrho_s > \varrho$ war. In der Frühphase des Weltalls bestimmte daher die Strahlungsdichte dessen Struktur. Der Zeitpunkt t_s markiert den Übergang (rückwärtsgehend) von einem durch ϱ bestimmten *Materiekosmos* zu einem durch ϱ_s bestimmten *Strahlungskosmos*. Aus den Beobachtungswerten ergibt sich $\varrho(t_s) = \varrho_s(t_s) \approx 5 \cdot 10^{-19}$ kg/m³ und $T(t_s) \approx 3\,000$ K. Nähert man sich immer weiter der kosmischen Singularität, steigt die Energiedichte der Strahlung (damit die Temperatur) immer weiter an. Sie wird so groß, daß Kernprozesse und schließlich auch Paarbildungsprozesse von Elementarteilchen stattfinden können.

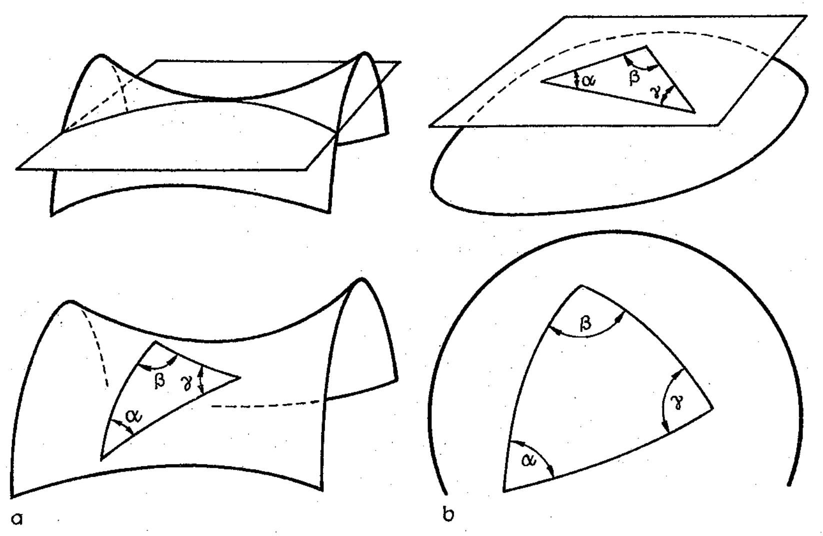


Abb. 12.4.2–1 Verschieden gekrümmte zweidimensionale Räume. Oben, dünne Linie: „flacher" (euklidischer) Raum mit der Krümmung Null, eine Ebene, in der für die Winkelsumme im Dreieck $\alpha + \beta + \gamma = 2\pi$ gilt. (a) hyperbolisch gekrümmter Raum (Krümmung < 0), Sattelfläche, auf der $\alpha + \beta + \gamma < 2\pi$ gilt; (b) sphärisch gekrümmter Raum (Krümmung > 0), Kugeloberfläche, auf der $\alpha + \beta + \gamma > 2\pi$ gilt

Geht man von den beobachteten Werten für ϱ_0 und H_0 aus, so ergibt sich, daß der Raum eine negative Krümmung hat. Dies bedeutet, daß die Expansion des Weltalls unbeschränkt andauert. Zwar wird sie verlangsamt – die gegenseitige Anziehung der kosmischen Massen bewirkt eine Verzögerung, die um so größer ist, je höher ϱ_0 ist –, die beobachtete mittlere Dichte ist aber zu gering, um die Expansionsgeschwindigkeit jemals auf Null abzubremsen.

Die Expansion würde dann aufhören und in eine Kontraktion übergehen, wenn die Materiedichte um einen Faktor 10 bis 20 größer als beobachtet wäre. Es ist aber unwahrscheinlich, daß eine so große Menge nichtleuchtender kosmischer Objekte sich der Beobachtung entzieht. Falls aber die Ruhmasse der im Weltall vorhandenen Neutrinos (vgl. 12.4.2.3.) nicht verschwindend klein ist, könnte die mittlere Gesamtmassendichte größer als die kritische sein, bei der die jetzige Expansion immer langsamer wird und in unendlich langer Zeit ganz verschwindet. Die Neutrinomasse (falls überhaupt vorhanden) ist aber bisher noch nicht angebbar.

Berücksichtigt man die gravitative Verzögerung der Expansion, so war früher die Ausdehnungsgeschwindigkeit größer als gegenwärtig, was bedeutet, daß die kosmische Singularität eine kürzere Zeitspanne (*Friedmann-Zeit* T_F) zurückliegt, als die Hubble-Zeit angibt. Der genaue Wert von T_F ist durch ϱ_0 und H_0 bedingt.

12.4.2.3. Frühphasen des Weltalls. Verfolgt man von der Gegenwart t_0 ausgehend die Expansion des Weltalls zurück, so war $R(t)$ früher kleiner, …

Aus diesen Überlegungen heraus läßt sich ein Modell über die frühen Entwicklungsphasen des Weltalls ableiten, dessen Basis die wohlbekannten, experimentell und theoretisch gesicherten Kenntnisse von Elementarteilchen- und Kernphysik bilden.

Die physikalische Entwicklung läßt sich quantitativ gesichert ab dem Zeitpunkt verfolgen, zu dem die der Strahlungsdichte äquivalente Massendichte unter die sank, die in den Atomkernen herrscht. Dieser Zustand wurde $\approx 10^{-6}$ s nach der Singularität erreicht; die Temperatur lag dabei bei $\approx 10^{13}$ K. Der vorausgehende Zustand läßt sich gegenwärtig nur qualitativ beschreiben. Das liegt an der noch vorhandenen Unvollständigkeit der Theorie der Elementarteilchen. Es besteht die Hoffnung, mittels einer großen, einheitlichen Theorie, die die starken, schwachen und elektromagnetischen Wechselwirkungen in einheitlicher Weise darzustellen vermag, auch die noch früher liegenden Zustände des Weltalls theoretisch modellieren zu können. Insgesamt dürfte aber die gegenwärtige Unkenntnis der extrem frühen Phasen des Weltalls die Gültigkeit der jetzigen Aussagen

nicht zu sehr behindern, da das Weltall in den ersten Sekundenbruchteilen thermodynamische Gleichgewichtszustände durchlief, in denen alle möglicherweise vorher existierenden größeren lokalen Störungen ausgelöscht waren. Unmittelbar nach der kosmischen Singularität herrschte eine solche extreme Energiedichte, daß alle Elementarteilchen und ihre Antiteilchen aus hochenergetischen Photonen gebildet werden konnten. Zur Entstehung eines Teilchens der Ruhmasse m muß für die Temperatur gelten $kt \gtrsim mc^2$. Bei Temperaturen mit $T \approx 10^{12}$ K, die etwa bis 10^{-4} s nach der Singularität herrschten, konnten daher schwere Teilchen, wie Protonen (P), Neutronen (N) und Mesonen (zusammengefaßt als Hadronen bezeichnet), gebildet werden. Am Ende dieser *Hadronenära* (Abb. 12.4.2–2), als die Temperatur unter 10^{12} K sank, zerstrahlten die meisten schweren Teilchen mit ihren Antiteilchen; zur Neubildung reichten nun aber die vorhandenen Energien nicht mehr aus. Offenbar existierte aber zu diesem Zeitpunkt im Weltall keine absolute Symmetrie mit überall exakt gleicher Anzahl von Teilchen und Antiteilchen, da es sonst überhaupt keine Protonen mehr im Weltall geben würde. Der relative Überschuß von Materie zu Antimaterie betrug aber nur $\approx 10^{-9}$, wie sich aus dem beobachteten Verhältnis der Photonendichte in der kosmischen Hintergrundstrahlung zur Protonendichte im Weltall abschätzen läßt. Im Rahmen einer großen, einheitlichen Theorie der Elementarteilchen scheint ein solcher Überschuß verständlich, da sie eine spontane Brechung der Symmetrie beinhaltet.

Teilchen geringer Masse (Leptonen), wie Elektronen (e) und Müonen einschließlich ihrer Antiteilchen, waren im Gleichgewicht mit dem Strahlungsfeld bis ≈ 10 s nach der Singularität ($T \approx 6 \cdot 10^9$ K) *(Leptonenära)*. Sie übertrafen dabei zahlenmäßig die schweren Teilchen. Bei Reaktionen der übriggebliebenen Protonen mit Elektronen (bzw. Neutronen mit Positronen) entstanden neben Neutronen (bzw. Protonen) auch eine große Zahl Neutrinos. Ihre mittlere freie Weglänge im Weltall wurde unterhalb von $\approx 10^{10}$ K so groß, daß die Neutrinos den Kosmos als freie Teilchen, entkoppelt von der übrigen Materie, durchfluten konnten. Unterhalb von $\approx 6 \cdot 10^9$ K fiel durch Zerstrahlung die Anzahl der Leptonen stark ab. Neben geringen Resten materieller Teilchen beherrschten fortan Lichtquanten das Weltall.

In der folgenden *Kernreaktionsära*, die bis $\approx 10^3$ s nach der Singularität dauerte, hatten Temperatur und Dichte solche

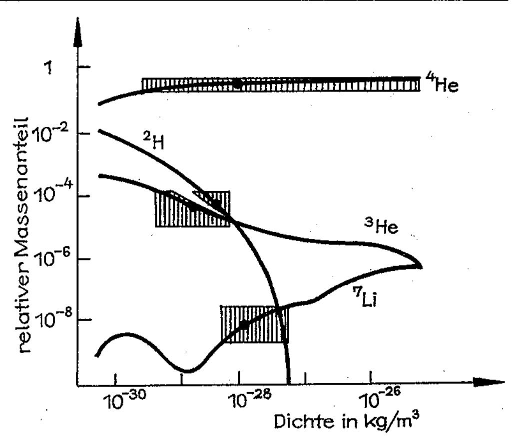


Abb. 12.4.2–3 Der relative Massenanteil von ^{2}H, ^{3}He, ^{4}He und ^{7}Li, der während der Kernreaktionsära gebildet wurde, in Abhängigkeit von der gegenwärtigen mittleren Dichte im Weltall. Die schraffierten Rechtecke geben den jeweiligen Unsicherheitsbereich für die beobachteten mittleren Häufigkeiten einzelner Kernsorten an. Die Daten legen eine mittlere Dichte von etwa 10^{-28} kg/m^3 im gegenwärtigen Weltall nahe

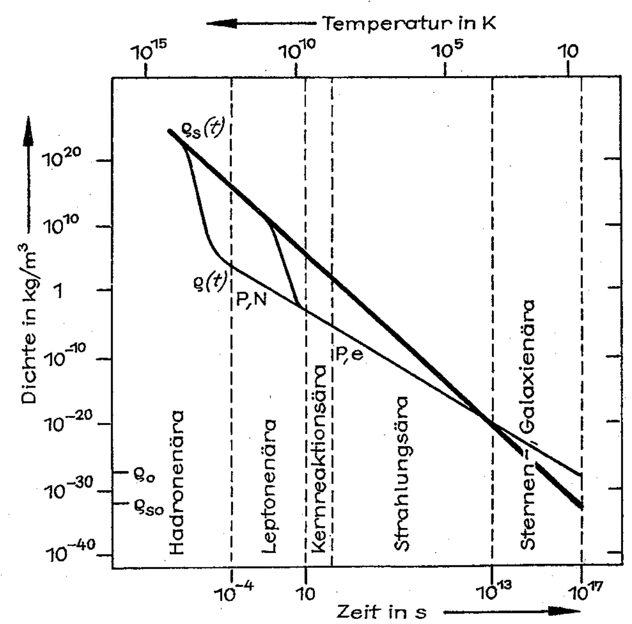


Abb. 12.4.2–2 Der Verlauf von Temperatur, Materiedichte ϱ und Strahlungsdichte ϱ_s in Abhängigkeit von der Zeit. ϱ_0 und ϱ_{s0} bezeichnen die gegenwärtige Materiedichte bzw. die zur Strahlungsdichte äquivalente Massendichte. Während der Leptonenära wird die Materiedichte wesentlich durch Protonen P und Neutronen N, während der Strahlungsära durch Protonen und Elektronen e festgelegt

Werte, daß die durch Kernreaktionen aus Protonen und Neutronen gebildeten Atomkerne nicht sofort wieder zerstört wurden. Aus dem zeitlichen Verlauf von Dichte und Temperatur läßt sich die Häufigkeit der sich bildenden und überlebenden Elemente berechnen. Dabei ergibt sich, daß neben Deuterium- nur noch Heliumkerne und eine sehr kleine Menge Lithiumkerne entstehen konnten, so daß am Ende der Kernreaktionsära ein Verhältnis von Protonen zu ^{4}He-Kernen wie $\approx$10:1 herrschte (Abb. 12.4.2.–3). Das gleiche Verhältnis erhält man bei der Bestimmung der gegenwärtigen mittleren kosmischen Elementenhäufigkeit (vgl. 12.2.2.4.). Der größte Teil des heute im Weltall vorhandenen Heliums wie auch Deuteriums dürfte daher aus der Frühzeit des Weltalls stammen.

Bei $\approx$3 000 K war die Kombination der Elektronen mit den Protonen abgeschlossen. Das Strahlungsfeld reichte nicht mehr zur Ionisation der Wasserstoffatome aus. Dabei entkoppelte sich die noch vorhandene Strahlung von der Materie, da die Wechselwirkungen fehlten. Die heute beobachteten Photonen der kosmischen Hintergrundstrahlung stammen aus dieser Zeit. Aus der nunmehr neutralen und von der Strahlung entkoppelten Materie konnten sich unter dem Einfluß der Gravitation aus anfangs kleinen Dichtefluktuationen größere Materieansammlungen, d. h. Sterne und Sternsysteme, bilden. Es begann die *Sternen-* und *Galaxienära*, die noch heute andauert. Die Frage nach dem Ursprung der ersten Dichtefluktuationen ist noch nicht beantwortet. Dies ist auch im Rahmen der hier dargestellten Modelle nicht möglich, da diese ja gerade Homogenität und Isotropie voraussetzen.

13. Ausgewählte mathematische Grundlagen und Methoden der Physik

Die Grundgesetze aller Teilgebiete der Physik sind mathematische Gleichungen. Überall dort, wo man nicht nur qualitativ veranschaulichen will, sondern die Grundgleichungen erläutert und aus diesen konkrete Aussagen ableitet, bedient man sich der Sprache der Mathematik. Dem Charakter der physikalischen Größen und der Form der Grundgleichungen entsprechend stehen dabei die Gebiete Algebra, Vektor- und Tensorrechnung, Differential- und Integralrechnung, gewöhnliche und partielle Differentialgleichungen und Vektoranalysis im Vordergrund. Daher werden diese Kapitel der Mathematik in dieser Darstellung vorrangig behandelt. Zu den mathematischen Begriffen und Aufgabenstellungen, die in den einzelnen Gebieten der Physik bei der Diskussion der Grundgesetze und deren Lösungen auftreten, können daher in diesem Teil Erläuterungen gefunden werden. Darüber hinaus sind die zahlreichen Beispiele in diesem Teil konkreten Fragestellungen der physikalischen Teilgebiete angepaßt.

13.1. Algebraische Strukturen

13.1.1. Gruppe

13.1.1.1. Gruppoid heißt eine Menge G zusammen mit einer (z. B. durch das Symbol $\circ$ bezeichneten) Vorschrift, die einem beliebigen geordneten Paar (g_1, g_2) von Elementen g_1, g_2 aus G eindeutig ein Element $g_3 \in G$ zuordnet: $g_3 = g_1 \circ g_2$ ($\in$ ist das Symbol für die Zugehörigkeit eines Elementes zu einer Menge).

Bsp. 13.1.1–1. Die Menge der Vektoren $\boldsymbol{a}$, $\boldsymbol{b}$, ... mit dem Vektorprodukt als Verknüpfungsvorschrift (vgl. 13.2.5.1.): $\boldsymbol{a} \circ \boldsymbol{b} = \boldsymbol{a} \times \boldsymbol{b}$ ist ein Gruppoid.

Bsp. 13.1.1–2. Die Menge der hermiteschen Matrizen (vgl. 13.2.1.2.; n Zeilen, n Spalten, n ist fest) mit der Matrizenmultiplikation (vgl. 13.2.1.1.) als Verknüpfung ist kein Gruppoid, da i. allg. das Produkt hermitescher Matrizen nicht hermitesch ist.

13.1.1.2. Halbgruppe heißt ein Gruppoid G, wenn für je drei Elemente aus G das *Assoziativgesetz* $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$ gilt. Das Bsp. 13.1.1–1 ist keine Halbgruppe (vgl. 13.2.5.1., Zerlegungssatz).

Bsp. 13.1.1–3. Die Menge G der ganzen positiven geraden Zahlen $2n$ ($n = 1, 2, \ldots$) mit der gewöhnlichen Zahlenmultiplikation als Verknüpfung ist eine Halbgruppe.

13.1.1.3. Halbgruppe mit Einselement. Eine Halbgruppe G mit *Einselement* $e \in G$ liegt vor, wenn für beliebige $g \in G$ stets $g \circ e = e \circ g = g$ gilt. Im Gegensatz zu Bsp. 13.1.1–3 besitzt Bsp. 13.1.1–4 ein Einselement.

Bsp. 13.1.1–4. Menge der ganzen positiven Zahlen 1, 2, ... mit der Multiplikation als Verknüpfung.

13.1.1.4. Gruppe ist eine Halbgruppe G mit Einselement $e \in G$, wobei zu jedem $g \in G$ ein *inverses* (*reziprokes*) Element $g^{-1} \in G$ existiert mit $g^{-1} \circ g = g \circ g^{-1} = e$. Eine Folgerung ist $(g_1 \circ g_2)^{-1} = g_2^{-1} \circ g_1^{-1}$ (Faktorreihenfolge!). Der Bruch g_1/g_2 ist i. allg. nicht bildbar, da i. allg. $g_2^{-1} \circ g_1 \neq g_1 \circ g_2^{-1}$ ist. Das Bsp. 13.1.1–4 ist keine Gruppe.

Bsp. 13.1.1–5. Die Menge der quadratischen Matrizen (n Zeilen, n Spalten, n ist fest), deren Determinanten (vgl. 13.2.1.1.) stets ungleich 0 sind, bildet mit der Matrizenmultiplikation als Verknüpfung eine Gruppe.

13.1.1.5. Abelsche Gruppe (*kommutative Gruppe*) heißt eine Gruppe G, bei der für beliebige g_1, g_2 aus G das *Kommutativgesetz* $g_1 \circ g_2 = g_2 \circ g_1$ gilt. Das Bsp. 13.1.1–5 ist keine Abelsche Gruppe. Anstatt in Abelschen Gruppen die Verknüpfung durch $\circ$ zu kennzeichnen, benutzt man meist das Pluszeichen $+$. Das Einselement e heißt dann *Nullelement* (Schreibweise 0), das reziproke Element von g wird dann durch $-g$ bezeichnet und heißt *entgegengesetztes* Element. Für $g + (-g) = -g + g = 0$ schreibt man kurz $g - g = -g + g = 0$.

13.1.2. Ring

Ein *Ring* ist eine Abelsche Gruppe G (Verknüpfungszeichen $+$), die bezüglich einer weiteren (multiplikativen) Verknüpfung $\circ$ ein Gruppoid ist, wobei beide Operationen durch die *Distributivgesetze* $g_3 \circ (g_1 + g_2) = g_3 \circ g_1 + g_3 \circ g_2$, $(g_1 + g_2) \circ g_3 = g_1 \circ g_3 + g_2 \circ g_3$ verbunden sind. Es ist oft üblich, das Zeichen $\circ$ wegzulassen.

Bsp. 13.1.2–1. Die additive Gruppe der Vektoren mit dem Vektorprodukt als weiterer Verknüpfung ist ein Ring.

13.1.2.1. Assoziativer Ring heißt ein Ring G, wenn das multiplikative Gruppoid G eine Halbgruppe ist, falls also $(g_1 g_2) g_3 = g_1 (g_2 g_3)$ gilt. Das Bsp. 13.1.2–1 ist kein assoziativer Ring (vgl. 13.2.5.1., Zerlegungssatz).

Bsp. 13.1.2–2. Die Menge der quadratischen Matrizen (n Zeilen, n Spalten, n ist fest) mit Matrizenaddition und -multiplikation bildet einen assoziativen Ring.

Bsp. 13.1.2–3. Läßt man im Bsp. 13.1.2–2 nur hermitesche Matrizen zu, so geht gemäß Bsp. 13.1.1–2 die Ringeigenschaft verloren.

13.1.2.2. Liescher Ring heißt ein Ring G, dessen multiplikative Verknüpfung (Bezeichnung $g_3 = [g_1, g_2]$) den Gleichungen $[g_1, g_2] = -[g_2, g_1]$ und $[g_1, [g_2, g_3]] + [g_2, [g_3, g_1]] + [g_3, [g_1, g_2]] = 0$ (*Jacobische Identität*) genügt. Das Bsp. 13.1.2–1 ist ein Liescher Ring.

Bsp. 13.1.2–4. Die Menge der hermiteschen Matrizen $\boldsymbol{A}$, $\boldsymbol{B}$, ... (n Zeilen, n Spalten, n ist fest) bildet einen Lieschen Ring, wenn neben der Matrizenaddition als Multiplikation die folgende *Kommutatorbildung* genommen wird: $[\boldsymbol{A}, \boldsymbol{B}] = c\mathrm{i}(\boldsymbol{AB} - \boldsymbol{BA})$, ($c$ ist feste reelle Zahl, $c \neq 0$, $\mathrm{i}^2 = -1$).

Bsp. 13.1.2–5. In der klassischen Mechanik bilden in einem gewissen Bereich alle dort unendlich oft partiell differenzierbaren Funktionen der verallgemeinerten Koordinaten q_k und der verallgemeinerten Impulse p_k ($k = 1, \ldots, f$; f = Zahl der Freiheitsgrade) bezüglich der gewöhnlichen Zahlenaddition und -multiplikation einen assoziativen Ring. Mit der Zahlenaddition und der durch die *Poisson-Klammer* $[g_1, g_2] = \sum_{k=1}^{f} \left(\frac{\partial g_1}{\partial q_k} \frac{\partial g_2}{\partial p_k} - \frac{\partial g_1}{\partial p_k} \frac{\partial g_2}{\partial q_k} \right)$

gegebenen Multiplikation liegt ein Liescher Ring vor.

13.1.3. Körper

Körper ist ein assoziativer Ring G, dessen multiplikative Halbgruppe G nach Entfernen des Nullelementes eine Abelsche Gruppe ergibt, d.h., daß zu jedem $g \in G$ mit $g \neq 0$ stets g^{-1} existiert und für alle $g_1, g_2 \in G$ stets $g_1 g_2 = g_2 g_1$ gilt.

Bsp. 13.1.3-1. Alle Zahlen der Gestalt $a + b\sqrt{2}$ (a, b sind rationale Zahlen) bilden einen Körper. So ist z.B.

$$\frac{a_1 + b_1\sqrt{2}}{a_2 + b_2\sqrt{2}} = \frac{(a_1 + b_1\sqrt{2})(a_2 - b_2\sqrt{2})}{(a_2 + b_2\sqrt{2})(a_2 - b_2\sqrt{2})}$$

$$= \frac{a_1 a_2 - 2b_1 b_2}{a_2^2 - 2b_2^2} + \frac{b_1 a_2 - a_1 b_2}{a_2^2 - 2b_2^2}\sqrt{2}.$$

13.1.3.1. Rationale Zahlen werden durch Brüche p/q dargestellt mit $q \neq 0$, wobei p und q ganze Zahlen sind. Die Menge der rationalen Zahlen bildet relativ zur gewöhnlichen Zahlenaddition und -multiplikation einen Körper. In der Menge der rationalen Zahlen ist eine Ordnungsbeziehung erklärt. Von zwei beliebigen rationalen Zahlen r_1 und r_2 gilt genau eine der drei Relationen: $r_1 < r_2$, $r_1 = r_2$, $r_2 < r_1$. Die *Ungleichheit* ist folgendermaßen mit der Addition und Multiplikation verknüpft: Aus $r_1 < r_2$ folgt für beliebige rationale Zahlen r stets $r_1 + r < r_2 + r$. Aus $r_1 < r_2$ folgt für alle $r > 0$ stets $r_1 r < r_2 r$.

13.1.3.2. Reelle Zahlen. Stellt man die rationalen Zahlen durch Punkte auf der Zahlengeraden dar, so bleiben noch Lücken, denen *irrationale Zahlen* entsprechen. Die Erweiterung der Menge der rationalen Zahlen durch Hinzunehmen der irrationalen Zahlen zur Menge der *reellen Zahlen* kann z.B. durch Intervallschachtelungen präzisiert werden. Die reellen Zahlen sind eineindeutig den Punkten der Zahlengeraden zugeordnet. Die reellen Zahlen bilden einen Körper und erfüllen die Gesetze über Ungleichheiten aus 13.1.3.1.

13.1.3.3. Komplexe Zahlen z werden eineindeutig den Punkten oder auch den Ortsvektoren der *Gaußschen Zahlenebene* (Abb. 13.1.3-1) zugeordnet: $z = x + iy = |z|\,e^{i\varphi}$ mit $|z| = (x^2 + y^2)^{1/2}$ und der *Eulerschen Formel* $e^{i\varphi} = \cos\varphi + i\sin\varphi$. Hierbei ist $|z|$ der *absolute Betrag* von z; dieser ist gleich dem Abstand des Punktes z vom Punkt 0. Mit φ wird ein orientierter Winkel bezeichnet; er wird gemessen von der positiven Richtung der Achse des Reellen aus bis zur Richtung des Ortsvektors z. Die reellen Zahlen x und y werden $x = \operatorname{Re} z$ = *Realteil* von z und $y = \operatorname{Im} z$ = *Imaginärteil* von z genannt. $z^* = x - iy$ heißt die *konjugiert komplexe Zahl* von $z = x + iy$. Man addiert, subtrahiert, multipliziert und dividiert komplexe Zahlen wie bei gewöhnlichen Zahlenrechnungen und setzt stets $i^2 = -1$. Also gilt mit $z_k = x_k + iy_k = |z_k| \exp(i\varphi_k)$, ($\exp a = e^a$; $k = 1, 2$) stets $z_1 + z_2 = x_1 + x_2 + i(y_1 + y_2)$, $z_1 z_2 = x_1 x_2 - y_1 y_2 + i(x_1 y_2 + x_2 y_1) = |z_1|\,|z_2| \exp(i(\varphi_1 + \varphi_2))$, $z_1/z_2 = z_1 z_2^*/(z_2 z_2^*)$

$$= \frac{x_1 x_2 + y_1 y_2}{x_2^2 + y_2^2} + \frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}\,i = \frac{|z_1|}{|z_2|}\exp(i(\varphi_1 - \varphi_2)).$$

Die Menge der komplexen Zahlen ist mit diesen Rechenoperationen ein Körper. Es ist ratsam, zwischen zwei komplexe Zahlen z_1 und z_2 nur dann ein Ungleichheitszeichen zu setzen, wenn sowohl z_1 als auch z_2 eine reelle Zahl ist, wenn also $\operatorname{Im} z_1 = \operatorname{Im} z_2 = 0$ ist. Es ist nämlich nicht möglich, zwischen komplexen Zahlen Ungleichheitsbeziehungen derart zu definieren, daß die Regeln für das Rechnen mit Ungleichheiten aus 13.1.3.1. übertragen werden können. Ist die quadratische Gleichung $x^2 + px + q = 0$ (p, q sind reell) zu lösen, so liefert die Lösungsformel $x_{1,2} = -(p/2) \pm [(p/2)^2 - q]^{1/2}$ im Fall $(p/2)^2 - q \geqq 0$ reelle Lösungen. Ist jedoch $(p/2)^2 - q < 0$, so ist in der Lösungsformel die Wurzel durch $i[q - (p/2)^2]^{1/2}$ zu ersetzen.

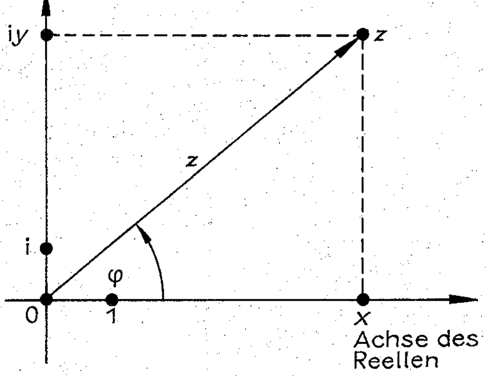


Abb. 13.1.3-1 Darstellung der komplexen Zahlen z in der Gaußschen Zahlenebene

13.1.4. Vektorräume

Vektorraum oder *linearer Raum* bezüglich des Körpers der reellen bzw. komplexen Zahlen (reeller bzw. komplexer Vektorraum) heißt eine Abelsche Gruppe G (Verknüpfungszeichen +), falls zu jedem Element $g, h, \ldots$ aus G und jeder reellen bzw. komplexen Zahl $c, c_1, c_2, \ldots$ ein Produkt $cg \in G$ erklärt ist und die folgenden Gleichungen gelten: $c(g + h) = cg + ch$, $(c_1 + c_2)g = c_1 g + c_2 g$, $(c_1 c_2)g = c_1(c_2 g)$, $1 \cdot g = g$.

Bsp. 13.1.4-1. Die gewöhnlichen Vektoren bilden mit der Vektoraddition und der Multiplikation von Vektoren mit reellen Zahlen einen reellen Vektorraum.

Bsp. 13.1.4-2. Die Menge der einspaltigen Matrizen mit n Zeilen (n ist fest), deren Elemente aus reellen bzw. komplexen Zahlen bestehen, bilden mit der Matrizenaddition und der Multiplikation von Matrizen mit reellen bzw. komplexen Zahlen einen rellen bzw. komplexen Vektorraum.

Bsp. 13.1.4-3. Die Menge $C[a, b]$ der in einem Intervall $a \leqq x \leqq b$ stetigen reellen bzw. komplexen Funktionen $y(x)$ der reellen unabhängigen Variablen x bilden bezüglich der gewöhnlichen Zahlenaddition und -multiplikation einen reellen bzw. komplexen Vektorraum.

Lineare Unabhängigkeit. n Elemente $\gamma_1, \ldots, \gamma_n$ aus dem Vektorraum G heißen *linear unabhängig*, wenn $\sum_{k=1}^{n} c_k \gamma_k = 0$ nur für $c_k = 0$ ($k = 1, \ldots, n$) richtig ist. Andernfalls heißen $\gamma_1, \ldots, \gamma_n$ *linear abhängig*. Ein unendliches System von Elementen aus G heißt linear unabhängig, wenn beliebig herausgegriffene endlich viele verschiedene Elemente des Systems linear unabhängig sind.

Bsp. 13.1.4-4. Im Bsp. 13.1.4-1 sind stets vier Vektoren voneinander linear abhängig. Drei Vektoren mit gemeinsamen Angriffspunkten sind genau dann linear unabhängig, wenn sie nicht in einer Ebene liegen. Im Bsp. 13.1.4-2 sind

je $n+1$ Matrizen voneinander linear abhängig. Im Bsp. 13.1.4–3 sind die unendlich vielen Funktionen x^k ($k=0, 1, 2, \ldots$) voneinander linear unabhängig.

Endliche Dimension n. Sind in einem Vektorraum stets $n+1$ Elemente voneinander linear abhängig, gibt es jedoch ein linear unabhängiges System von n Elementen: $\gamma_1, \ldots, \gamma_n$, so heißt dieses *Basis des n-dimensionalen Vektorraumes G*. Jedes Element $g \in G$ ist dann auf genau eine Weise als *Linearkombination* der Basiselemente darstellbar:

$$g = \sum_{k=1}^{n} c_k \gamma_k .$$

Bsp. 13.1.4–5. Die Vektorräume der Bsp. 13.1.4–1 bzw. 13.1.4–2 sind drei- bzw. n-dimensional. Das Bsp. 13.1.4–3 ist ein unendlich dimensionaler Vektorraum.

13.1.4.1. Hilbert-Räume. Skalarprodukt. Ordnet man in einem reellen bzw. komplexen Vektorraum G jedem geordneten Paar (g, h) von Elementen g, h aus G eine reelle bzw. komplexe Zahl $\langle g \mid h \rangle$ derart zu, daß $\langle g \mid h \rangle = \langle h \mid g \rangle^*$, $\langle g \mid c_1 h_1 + c_2 h_2 \rangle = c_1 \langle g \mid h_1 \rangle + c_2 \langle g \mid h_2 \rangle$, $\langle g \mid g \rangle \geqq 0$ und „$\langle g \mid g \rangle = 0$ nur für $g = 0$" gilt, so heißt $\langle g \mid h \rangle$ *Skalarprodukt* und G *Prä-Hilbert-* oder *unitärer Raum*. Ist $\langle g \mid h \rangle = 0$, so heißen g und h zueinander *orthogonal*.

Bsp. 13.1.4–6. Im Bsp. 13.1.4–1 wird üblicherweise das Skalarprodukt durch $\langle \boldsymbol{a} \mid \boldsymbol{b} \rangle = \boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}| \cos \varphi$ eingeführt, wobei $\varphi (0 \leqq \varphi \leqq \pi)$ der Winkel zwischen $\boldsymbol{a}$ und $\boldsymbol{b}$ ist. Im Bsp. 13.1.4–2 kann man mit der gewöhnlichen Matrizenmultiplikation (vgl. 13.2.1.1.) das Skalarprodukt durch $\langle \boldsymbol{a} \mid \boldsymbol{b} \rangle = (\boldsymbol{a}^*)^{\mathrm{T}} \boldsymbol{b}$ einführen. Im Bsp. 13.1.4–3 ergibt sich mit der gegebenen *Gewichtsfunktion* $p(x)$ [$p(x)$ fest, reell, stetig, $p(x) \geqq 0$, $p(x)$ besitzt in $a \leqq x \leqq b$ höchstens endliche viele Nullstellen] ein Skalarprodukt

durch $\langle y_1(x) \mid y_2(x) \rangle = \int\limits_a^b p(x)\,(y_1(x))^* y_2(x)\,\mathrm{d}x$.

Norm. Angeregt durch den Zusammenhang zwischen dem absoluten Betrag und dem Skalarprodukt in der Vektorrechnung ($|\boldsymbol{a}| = \sqrt{\boldsymbol{a} \cdot \boldsymbol{a}}$) wird die *Norm* $\|g\|$ eines Elementes g aus dem Prä-Hilbert-Raum G durch $\|g\| = \sqrt{\langle g \mid g \rangle}$ eingeführt. Dann gilt: $\|g+h\| \leqq \|g\| + \|h\|$, $\|cg\| = |c|\,\|g\|$, „$\|g\| = 0$ nur für $g = 0$".

Metrik. Unter der Entfernung der beiden Elemente g und h aus G versteht man $\varrho(g, h) = \|g - h\|$. Sie hat die Eigenschaften $\varrho(g, h) = \varrho(h, g)$, $\varrho(g, h) \leqq \varrho(g, h_1) + \varrho(h_1, h)$ (*Dreiecksungleichung*), $\varrho(g, h) \geqq 0$, „$\varrho(g, h) = 0$ nur für $g = h$".

Metrischer Raum. Eine Teilmenge M eines Prä-Hilbert-Raumes G ($M \subseteqq G$) ist im allgemeinen selbst kein Prä-Hilbert-Raum, weil die Addition oder auch die Multiplikation mit Zahlen aus M hinausführen kann. M ist jedoch ein *metrischer Raum*, wo zwischen je zwei Elementen g und h aus M eine Entfernung $\varrho(g, h)$ definiert ist, die den oben genannten Relationen genügt.

Konvergenz. Man sagt, daß in einem Prä-Hilbert-Raum G (bzw. metrischen Raum M) eine Folge von Elementen $u_1, u_2, \ldots \in G$ (bzw. $\in M$) zum Grenzwert $A \in G$ (bzw. $A \in M$) strebt – man schreibt in diesem Fall $\lim\limits_{n \to \infty} u_n = A$ oder auch $u_n \to A$ für $n \to \infty$ –, wenn der Abstand $\varrho(u_n, A)$ zwischen u_n und A beliebig klein wird für alle hinreichend großen n.

Häufungspunkt. Das Element A aus dem Prä-Hilbert-Raum G heißt *Häufungspunkt* der Teilmenge M von G, falls es in M eine Folge $u_n \neq A$ ($n = 1, 2, \ldots$) mit $\lim\limits_{n \to \infty} u_n = A$ gibt.

Abgeschlossene Hülle. Nimmt man zu einer Teilmenge M des Prä-Hilbert-Raumes G alle Häufungspunkte von M (relativ zu G) hinzu, so entsteht die *abgeschlossene Hülle* $\overline{M}$ von M.

Lineare Hülle. Bildet man von einer Teilmenge M eines reellen bzw. komplexen Vektorraumes G die Menge $L(M)$ aller Linearkombinationen $\sum\limits_{k=1}^{n} c_k u_k$ (n ist nicht fest, $u_k \in M$, c_k beliebige reelle bzw. komplexe Zahlen), so ist $L(M)$ ein reeller bzw. komplexer Vektorraum, den man die *lineare Hülle* von M nennt.

Basis im unendlich-dimensionalen Hilbert-Raum. Man sagt, das System der unendlich vielen voneinander linear unabhängigen Elemente $\gamma_1, \gamma_2, \ldots$ des Prä-Hilbert-Raumes G sei eine *Basis*, wenn die lineare Hülle $L(\gamma_1, \gamma_2, \ldots)$ dieser Elemente durch den Prozeß der Abschließung den Prä-Hilbert-Raum G ergibt: $\overline{L(\gamma_1, \gamma_2, \ldots)} = G$.

Orthogonale Basis. Die Basis $\varphi_1, \varphi_2, \ldots$ des Prä-Hilbert-Raumes heißt orthogonal, wenn $\langle \varphi_k \mid \varphi_l \rangle = 0$ für $k \neq l$ gilt.

Orthonormale Basis. Eine orthogonale Basis $\Psi_1, \Psi_2, \ldots$ ist sogar orthonormal, falls $\|\Psi_k\| = 1$ ist.

Orthonormalisierungsverfahren. Erhard *Schmidt* zeigte, wie man aus einer Basis $\gamma_1, \gamma_2 \ldots$ eine orthonormale Basis $\Psi_1, \Psi_2, \ldots$ herstellen kann: $\Psi_1 = \|\gamma_1\|^{-1} \gamma_1$, $\Psi_k = \|h_k\|^{-1} h_k$ mit $h_k = \gamma_k - \sum\limits_{l=1}^{k-1} \langle \gamma_k \mid \Psi_l \rangle \Psi_l$ $(k = 2, 3, \ldots)$.

Fourier-Entwicklung. Relativ zu einer orthogonalen Basis $\varphi_1, \varphi_2, \ldots$ des Prä-Hilbert-Raumes G kann jedes $g \in G$ durch eine *Fouriersche Reihe* $g = \lim\limits_{n \to \infty} \sum\limits_{k=1}^{n} c_k \varphi_k = \sum\limits_{k=1}^{\infty} c_k \varphi_k$ dargestellt werden, wobei für die *Fourier-Koeffizienten* $c_k = \|\varphi_k\|^{-2} \langle \varphi_k \mid g \rangle$ gilt.

Bsp. 13.1.4–7. Im Bsp. 13.1.4–1 ist $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ eine orthonormale Basis, wobei $\boldsymbol{e}_k$ ($k = 1, 2, 3$) ein Vektor mit dem absoluten Betrag $|\boldsymbol{e}_k| = 1$ ist und bei vorliegendem kartesischen Koordinatensystem (x_1, x_2, x_3) in Richtung der x_k-Achse weist. Im Bsp. 13.1.4–2 ist mit dem Skalarprodukt $\langle \boldsymbol{a} \mid \boldsymbol{b} \rangle = (\boldsymbol{a}^*)^{\mathrm{T}} \boldsymbol{b}$ ein System von n voneinander verschiedenen einspaltigen Matrizen eine orthonormale Basis, falls jede dieser Matrizen genau ein Element besitzt, das gleich 1 ist, während ihre anderen Elemente alle gleich 0 sind. Für das Bsp. 13.1.4–3 in Verbindung mit dem Bsp. 13.1.4–6 werden hier einige orthogonale Basen mit zugehöriger *Gewichtsfunktion* $p(x)$ und spezieller Wahl von $a \leqq x \leqq b$ angegeben.

Fourier-Cosinus-Sinus-Entwicklung: $1/2$, $\cos \dfrac{2\pi\nu x}{l}$, $\sin \dfrac{2\pi\nu x}{l}$

$(\nu = 1, 2, \ldots)$; $a \leqq x \leqq a + l$, $\langle u \mid v \rangle = \int\limits_a^{a+l} u^* v\,\mathrm{d}x$;

Norm: $\|1/2\| = l^{(1/2)}/2$, $\left\| \cos \dfrac{2\pi\nu x}{l} \right\| = \left\| \sin \dfrac{2\pi\nu x}{l} \right\| = (l/2)^{1/2}$.

Fourier-Exponentialentwicklung: $\exp\left(\dfrac{2\pi\nu \mathrm{i} x}{l}\right)$ $(\nu = 0, \pm 1, \pm 2, \ldots;\ \mathrm{i}^2 = -1)$; Norm: $\left\| \exp\left(\dfrac{2\pi\nu \mathrm{i} x}{l}\right) \right\| = \sqrt{l}$.

Legendresche Polynome: $P_n(x) = \dfrac{1}{2^n n!} \dfrac{\mathrm{d}^n}{\mathrm{d}x^n} [(x^2 - 1)^n]$

$(n = 0, 1, 2, \ldots)$; $-1 \leqq x \leqq 1$, $\langle u \mid v \rangle = \int\limits_{-1}^{+1} u^* v\,\mathrm{d}x$;

Norm: $\|P_n(x)\| = \sqrt{2/(2n+1)}$.

Hermitesche Polynome: $H_n(x) = (-1)^n \exp(x^2) \mathrm{d}^n/\mathrm{d}x^n [\exp(-x^2)]$ $(n = 0, 1, 2, \ldots)$; $-\infty < x < +\infty$, $\langle u \mid v \rangle = \int_{-\infty}^{+\infty} \exp(-x^2) u^* v \, \mathrm{d}x$; Norm: $\|H_n(x)\| = \sqrt{2^n n!}\,\sqrt[4]{\pi}$.

Laguerresche Polynome: $L_n(x) = (\mathrm{e}^x/n!)(\mathrm{d}^n/\mathrm{d}x^n)(x^n \mathrm{e}^{-x}) = \sum_{\nu=0}^{n} \frac{(-1)^\nu}{\nu!}\binom{n}{\nu} x^\nu$ $(n = 0, 1, 2, \ldots)$; $0 \leqq x < \infty$,

$\langle u \mid v \rangle = \int_0^\infty \mathrm{e}^{-x} u^* v \, \mathrm{d}x$; Norm: $\|L_n(x)\| = 1$.

Bsp. 13.1.4–8. Die Menge der auf der Kugeloberfläche $x^2 + y^2 + z^2 = 1$ stetigen Funktionen $u(\vartheta, \varphi)$ [(r, ϑ, φ): Kugelkoordinaten (vgl. Bsp. 13.2.5–2), $r = 1$, $0 \leqq \vartheta \leqq \pi$, $0 \leqq \varphi \leqq 2\pi$] bilden mit dem Skalarprodukt

$$\langle u \mid v \rangle = \int_0^{2\pi} \left\{ \int_0^\pi u^*(\vartheta, \varphi)\, v(\vartheta, \varphi) \sin\vartheta \,\mathrm{d}\vartheta \right\} \mathrm{d}\varphi$$

einen Prä-Hilbert-Raum. Als Basis können die normierten *Grundfunktionen der Kugelflächenfunktionen* genommen werden (k bei Y ist kein Exponent, sondern ein *oberer Index*):

$$Y_n^k(\vartheta, \varphi) = \sqrt{\left[\frac{(2n+1)}{4\pi}\right]\left[\frac{(n-|k|)!}{(n+|k|)!}\right]}\,(1-x^2)^{|k|/2} \frac{\mathrm{d}^{|k|}}{\mathrm{d}x^{|k|}}$$

$P_n(x)\,\mathrm{e}^{\mathrm{i}k\varphi}$ [$x = \cos\vartheta$, $P_n(x)$ = Legendresches Polynom, $n = 0, 1, 2, \ldots$; $k = 0 \pm 1, \ldots, \pm n$], $\|Y_n^k(\vartheta, \varphi)\| = 1$.

Cauchy-Folge. In einem Prä-Hilbert-Raum G bzw. metrischen Raum M ist jede konvergente Folge $u_1, u_2, \ldots$ eine *Cauchy-Folge* (auch *Fundamentalfolge* oder auch *in sich konvergente Folge* genannt), d. h. der Abstand $\varrho(u_n, u_m)$ zweier Elemente u_n und u_m wird beliebig klein, wenn nur n und m hinreichend groß sind.

Hilbert-Raum. Ist im Prä-Hilbert-Raum G jede Cauchy-Folge $u_1, u_2, \ldots$ eine konvergente Folge, d. h. gilt $\lim_{n\to\infty} u_n = A$ mit $A \in G$, so heißt G Hilbert-Raum. Ist dies nicht der Fall, so kann man auf genau eine Weise den Prä-Hilbert-Raum G durch Hinzunahme neuer Elemente zu einem Hilbert-Raum vervollständigen. Liegt eine Basis im Prä-Hilbert-Raum vor, so geht diese Eigenschaft beim Übergang zum Hilbert-Raum nicht verloren.

Bsp. 13.1.4–9. Mit den Skalarprodukten aus Bsp. 13.1.4–6 sind die Bsp. 13.1.4–1 und 13.1.4–2 Hilbert-Räume. In dem Bsp. 13.1.4–3 mit Bsp. 13.1.4–7 dagegen müssen noch gewisse unstetige Funktionen hinzugenommen werden, um einen Hilbert-Raum zu erhalten.

13.1.4.2. Banach-Räume. Prä-Banach-Raum. Führt man in einem Vektorraum G nicht wie in 13.1.4.1. axiomatisch ein Skalarprodukt, sondern eine Norm $\|u\| \geqq 0$ mit den in 13.1.4.1. genannten Eigenschaften ein, so erhält man einen *Prä-Banach-Raum* (*normierten Raum*).

Metrik, Konvergenz, abgeschlossene Hülle, lineare Hülle, Basis, Cauchy-Folge. Diese Begriffe – also ausgenommen die Orthogonalität – können wörtlich aus 13.1.4.1. übernommen werden.

Banach-Raum. Ist im Prä-Banach-Raum jede Cauchy-Folge konvergent, so heißt er *Banach-Raum*. Es gibt Banach-Räume, die keine Prä-Hilbert-Räume sind, deren Norm also nicht durch ein geeignetes Skalarprodukt konstruierbar ist.

Bsp. 13.1.4–10. Der Raum $C[a, b]$ (vgl. Bsp. 13.1.4–3) mit der *Maximumnorm* $\|y\| = \max_{a \leqq x \leqq b} |y(x)|$ ist ein Banach-Raum, jedoch kein Prä-Hilbert-Raum.

13.1.4.3. Algebra über einem Körper. Ein reeller bzw. komplexer Vektorraum G, der gleichzeitig ein Ring ist, wobei die multiplikative Verknüpfung $\circ$ im Ring mit der Multiplikation mit reellen bzw. komplexen Zahlen c den Gleichungen $c(g \circ h) = (cg) \circ h = g \circ (ch)$ genügt, heißt reelle bzw. komplexe *Algebra*.

Bsp. 13.1.4–11. Nimmt man in den Bsp. 13.1.2–1, 13.1.2–2, 13.1.2–4, 13.1.2–5 die Multiplikation von reellen bzw. komplexen Zahlen mit den dortigen Elementen (Vektoren oder Matrizen oder Funktionen) hinzu, so entsteht eine Algebra, die im jeweiligen Beispiel den Zusatz Liesch oder assoziativ erhält.

13.1.5. Lineare Operatoren

G_1 und G_2 seien beide reelle bzw. komplexe Vektorräume. $D(T)$ sei ein Vektorraum, der (mit den Rechenoperationen aus G_1) eine Teilmenge von G_1 ist: $D(T) \subseteqq G_1$. (Eine Menge wird auch als Teilmenge von sich selbst aufgefaßt: $G_1 \subseteqq G_1$). *Linearer Operator* ist eine *Zuordnungsvorschrift*, die jedem Element u aus dem *Definitionsbereich* $D(T)$ genau ein Element w aus G_2 zuordnet [Schreibweise: $w = T(u)$ oder auch $w = Tu$], wobei für alle reellen bzw. komplexen Zahlen $c_1, c_2, \ldots$ und beliebige Elemente $u_1, u_2, \ldots$ aus $D(T)$ stets $T(c_1u_1 + c_2u_2) = c_1Tu_1 + c_2Tu_2$ gilt. Durchläuft u alle Elemente von $D(T)$, so durchläuft $w = T(u)$ einen Vektorraum, der Teilmenge von G_2 ist und Wertebereich von T genannt wird. Ist speziell G_2 der Körper der reellen bzw. komplexen Zahlen, so nennt man den linearen Operator auch *lineares Funktional*.

Bsp. 13.1.5–1. Läßt man in einer Algebra über dem Körper der reellen oder komplexen Zahlen (vgl. 13.1.4.3.) bei der multiplikativen Verknüpfung den einen Faktor fest, während der andere Faktor alle Elemente aus G durchläuft, so liegt ein linearer Operator vor, dessen Definitionsbereich mit G zusammenfällt und dessen Wertebereich in G enthalten ist.

Bsp. 13.1.5–2. Es sei $\varphi_1, \varphi_2, \ldots$ eine orthogonale Basis eines Prä-Hilbert-Raumes G. Ordnet man jedem Element $u \in G$ die Folge seiner Fourier-Koeffizienten $\|\varphi_k\|^{-2}\langle\varphi_k \mid u\rangle$ $(k = 1, 2, \ldots)$ zu, so ist damit ein linearer Operator gegeben. Sein Definitionsbereich ist G, sein Wertebereich ist in der Menge aller Zahlenfolgen enthalten, wobei die Addition von Zahlenfolgen und Multiplikation von Zahlenfolgen mit einer Zahl in üblicher Weise vorgenommen wird.

Bsp. 13.1.5–3. Eine Matrix vom Typ (m, n) kann als linearer Operator aufgefaßt werden, dessen Definitionsbereich eine Menge von Matrizen des Typs $(n, 1)$ ist und dessen Wertebereich in der Menge der Matrizen vom Typ $(m, 1)$ enthalten ist (vgl. Bsp. 13.1.4–2), wobei die Zuordnungsvorschrift mittels der gewöhnlichen Matrizenmultiplikation realisiert wird.

Bsp. 13.1.5–4. Läßt man im Prä-Hilbert-Raum G bei der Skalarproduktbildung $\langle g \mid h\rangle$ den ersten Faktor fest, während der zweite in G variiert, so ist damit ein lineares Funktional gegeben.

Bsp. 13.1.5–5. Durch $T = a_\nu(x)\,\mathrm{d}^\nu/\mathrm{d}x^\nu$ ist ein *linearer Differentialoperator* gegeben. Sein Definitionsbereich ist eine Teilmenge von $C[a, b]$ (vgl. Bsp. 13.1.4–3), sein Wertebereich ebenfalls, falls $a_\nu(x)$ stetig ist. Unter der Multiplikation von $\mathrm{d}^\nu/\mathrm{d}x^\nu$ mit $y = y(x)$ versteht man die ν-te-Ableitung von $y(x)$, also $\mathrm{d}^\nu y/\mathrm{d}x^\nu$. Ist $\nu = 0$, so ist definitionsgemäß die nullte Ableitung von y die Funktion y selbst.

Bsp. 13.1.5–6. *Fourier-Transformation*. Durch das *Fourier-Integral* $F(y) = (1/\sqrt{2\pi}) \int_{-\infty}^{+\infty} f(x)\,e^{ixy}\,dx$ ist ein linearer *Integraloperator* gegeben. Sowohl der Definitionsbereich als auch der Wertebereich besteht aus komplexen Funktionen einer reellen unabhängigen Variablen x bzw. y.

Bsp. 13.1.5–7. *Laplace-Transformation*. Sie ist ein linearer Integraloperator und wird durch $F(p) = \int_0^\infty e^{-pt} f(t)\,dt$ gegeben. Der Definitionsbereich besteht aus komplexen Funktionen $f(t)$ der reellen unabhängigen Variablen t, während der Wertebereich aus komplexen Funktionen $F(p)$ der komplexen unabhängigen Variablen p besteht.

13.1.5.1. Vektorraum linearer Operatoren. Die Menge M linearer Operatoren $T_1, T_2, \ldots$ mit gemeinsamen Definitionsbereich D und Wertebereichen, die alle im Vektorraum G_2 liegen, wird zu einem Vektorraum, wenn die *Linearkombination von Operatoren* $T = c_1T_1 + c_2T_2$ durch $Tu = c_1T_1u + c_2T_2u$ erklärt wird.

Bsp. 13.1.5–8. Mit Bsp. 13.1.5–5 ist $\sum_{\nu=0}^{n} a_\nu(x)\,d^\nu/dx^\nu$ ein linearer Differentialoperator.

13.1.5.2. Assoziative Algebra linearer Operatoren. Sind in 13.1.5.1. alle Wertebereiche von $T_1, T_2, \ldots$ sogar Teilmengen von D, so definiert man das Produkt $T = T_1T_2$ durch die *Hintereinanderausführung* $Tu = T_1(T_2[u])$, i. allg. ist $T_1T_2 \neq T_2T_1$. Damit wird der Vektorraum der linearen Operatoren aus 13.1.5.1. zu einer assoziativen Algebra (vgl. 13.1.4.3.).

Bsp. 13.1.5–9. Zwischen der zeitlichen Ableitung d/dt eines Vektors $\boldsymbol{a}$ in einem Inertialsystem und der zeitlichen Ableitung $\tilde{d}/dt$ von $\boldsymbol{a}$ in einem rotierenden Bezugssystem, das durch den konstanten Vektor der Winkelgeschwindigkeit $\boldsymbol{\omega}$ beschrieben wird, besteht der Zusammenhang $d/dt = \tilde{d}/dt + \boldsymbol{\omega}\times$. Folglich ergibt sich für die zweite zeitliche Ableitung $(d^2/dt^2)\,\boldsymbol{a} = (\tilde{d}/dt + \boldsymbol{\omega}\times)(\tilde{d}/dt + \boldsymbol{\omega}\times)\,\boldsymbol{a} = (\tilde{d}^2/dt^2)\boldsymbol{a} + 2\boldsymbol{\omega}\times(\tilde{d}/dt)\boldsymbol{a} + \boldsymbol{\omega}\times(\boldsymbol{\omega}\times\boldsymbol{a})$.

13.1.5.3. Adjungierter Operator. Im Hilbert-Raum G sei ein linearer Operator T gegeben, wobei die abgeschlossene Hülle des Definitionsbereiches $D(T)$ von T (vgl. 13.1.4.1.) gleich G ist: $\overline{D(T)} = G$. Der Wertebereich von T sei eine Teilmenge von G. Dann gibt es den zu T *adjungierten* linearen Operator T^+, d.h. es gilt für alle $u \in D(T) \subseteqq G$ und alle $v \in D(T^+) \subseteqq G$ stets $\langle Tu \mid v\rangle = \langle u \mid T^+v\rangle$.

Bsp. 13.1.5–10. Der Prä-Hilbert-Raum der stetigen Funktionen $C[0, l]$ (vgl. Bsp. 13.1.4–6, 13.1.4–3) mit dem Skalarprodukt

$$\langle u \mid v\rangle = \int_0^l u^*(x)v(x)\,dx$$

werde durch Hinzunahme weiterer Funktionen zu einem Hilbert-Raum G ergänzt (vgl. 13.1.4.1.). Die Menge der Funktionen $y = y(x)$, $(0 \leqq x \leqq l)$, die eine stetige Ableitung besitzen und der *linearen homogenen Randbedingung* $y(0) = cy(l)$ genügen (lineare homogene Randbedingungen liegen vor, wenn alle Funktionen, die diesen Bedingungen genügen, einen Vektorraum bilden), seien der Definitionsbereich $D(T)$ des linearen Operators $T = -i\,d/dx$. Die abgeschlossene Hülle von $D(T)$ ist gleich G: $\overline{D(T)} = G$. Zur Bestimmung des zugehörigen adjungierten Operators wird zunächst mittels partieller Integration notiert:

$$\begin{aligned}\langle Tu \mid v\rangle &= \int_0^l (-iu')^*v\,dx = i\int_0^l (u')^*v\,dx = i[u^*v]_0^l - i\int_0^l u^*v'\,dx\\ &= i[u^*(l)v(l) - u^*(0)v(0)] + \int_0^l u^*(-iv')\,dx\\ &= iu^*(l)[v(l) - c^*v(0)] + \langle u \mid Tv\rangle.\end{aligned}$$

Folglich lautet der adjungierte Differentialoperator ebenfalls $-i(d/dx)$, allerdings müssen die Funktionen seines Definitionsbereiches der *adjungierten Randbedingung* $y(l) = c^*y(0)$ genügen.

13.1.5.4. Lineare Operatoren mit speziellen Eigenschaften. Hermitesche (symmetrische) Operatoren. Gilt in 13.1.5.3. $D(T) \subseteqq D(T^+) \subseteqq G$ und stimmen T und T^+ für alle $u \in D(T) \subseteqq G$ überein, so heißt T *hermitesch* (*symmetrisch*). Es gilt dann also $\langle Tu \mid v\rangle = \langle u \mid Tv\rangle$ für alle $u \in D(T)$.

Bsp. 13.1.5–11. Ist im Bsp. 13.1.5–10 $|c| = 1$, so liegt dort ein *hermitescher Differentialoperator* vor.

Bsp. 13.1.5–12. Vgl. Bsp. 13.1.4–11 in Verbindung mit Bsp. 13.1.2–4, wobei die hermiteschen Matrizen durch hermitesche Operatoren ersetzt werden.

Unitäre Operatoren. Ist in 13.1.5.3. der Wertebereich von T sogar gleich G, so heißt T unitär, falls für alle u, v aus G stets $\langle Tu \mid Tv\rangle = \langle u \mid v\rangle$ gilt.

13.1.6. Boolesche Algebra

13.1.6.1. Verband. Eine nichtleere Menge V von Elementen $E_1, E_2, \ldots$ heißt ein *Verband*, wenn je zwei Elementen E_1, E_2 aus V die Elemente $E_1 \cap E_2$ (*Durchschnitt*) und $E_1 \cup E_2$ (*Vereinigung*) aus V zugeordnet sind, so daß die folgenden Gesetze gelten:

Kommutative Gesetze.
$E_1 \cap E_2 = E_2 \cap E_1$; $E_1 \cup E_2 = E_2 \cup E_1$.

Assoziative Gesetze.
$(E_1 \cap E_2) \cap E_3 = E_1 \cap (E_2 \cap E_3)$; $(E_1 \cup E_2) \cup E_3 = E_1 \cup (E_2 \cup E_3)$.

Verschmelzungsgesetze.
$E_1 \cap (E_1 \cup E_2) = E_1$; $E_1 \cup (E_1 \cap E_2) = E_1$.

Bsp. 13.1.6–1. Die Teilmengen einer festen Menge M bilden bezüglich des mengentheoretischen Durchschnittes und der mengentheoretischen Vereinigung einen Verband. $E_1 \cap E_2$ wird dann gelesen „E_1 und E_2“, $E_1 \cup E_2$ wird gelesen „E_1 oder E_2“.

13.1.6.2. Distributiver Verband. Ein Verband heißt distributiv, wenn in ihm das *distributive Gesetz* $E_1 \cup (E_2 \cap E_3) = (E_1 \cup E_2) \cap (E_1 \cup E_3)$ gilt. Hieraus folgt das weitere distributive Gesetz $E_1 \cap (E_2 \cup E_3) = (E_1 \cap E_2) \cup (E_1 \cap E_3)$.

13.1.6.3. Verband mit Null und Eins. Im *Verband* V mit *Null* und *Eins* existieren Elemente 0 und 1, wobei für jedes $E \in V$ immer $E \cap 0 = 0$ und $E \cup 1 = 1$ gilt.

13.1.6.4. Komplementärer Verband. In einem Verband V mit Null und Eins heißt $\overline{E} \in V$ *Komplement* von $E \in V$, wenn $E \cap \overline{E} = 0$ und $E \cup \overline{E} = 1$ gilt. Existiert zu jedem E mindestens ein $\overline{E}$, so spricht man vom *komplementären Verband*. $\overline{E}$ ist eindeutig festgelegt, wenn V darüber hinaus noch distributiv ist.

13.1.6.5. Boolesche Algebra (*Boolescher Verband*) ist ein distributiver und komplementärer Verband.

Bsp. 13.1.6–2. Entsprechend dem Sachverhalt, daß beim Ablauf von Versuchen das Resultat im Rahmen verschiedener Möglichkeiten ungewiß ist, ordnet man in der Wahrscheinlichkeitstheorie jedem derartigen Zufallsexperiment

eine Menge $\mathfrak{E}$ von Ereignissen $E_1, E_2, \ldots$ zu. Man setzt voraus, daß $\mathfrak{E}$ eine Boolesche Algebra ist. Die Boolesche Algebra heißt im jetzigen Zusammenhang *Ereignisalgebra* (*Ereignisfeld*). Das Nullelement 0 heißt hierbei *unmögliches Ereignis*, das Einselement 1 (Schreibweise hier häufig I) heißt *sicheres Ereignis*. Im Fall $E_1 \cap E_2 = 0$ heißen E_1 und E_2 *unvereinbar*. $E \in \mathfrak{E}$ heißt ein *atomares Ereignis*, wenn sich E nicht in der Gestalt $E = E_1 \cup E_2$ mit $E_1, E_2 \in \mathfrak{E}$, $E_1 \neq E$, $E_2 \neq E$ darstellen läßt.

13.1.6.6. Wahrscheinlichkeitsalgebra (*Wahrscheinlichkeitsfeld*) liegt vor, wenn jedem Element E der vorliegenden Ereignisalgebra $\mathfrak{E}$ (vgl. Bsp. 13.1.6–2) eine reelle Zahl $P(E)$ – die *Wahrscheinlichkeit* von E – so zugeordnet wird, daß die folgenden *Axiome von Kolmogorov* erfüllt sind: Aus $E \in \mathfrak{E}$ folgt $0 \leqq P(E) \leqq 1$; $P(I) = 1$ (*Normierungsaxiom*). Aus $E_k \in \mathfrak{E}$ und $E_k \cap E_l = 0$ ($k \neq l$; $k, l = 1, 2, \ldots$) folgt $\bigcup_{k=1}^{\infty} E_k \in \mathfrak{E}$ und $P\left(\bigcup_{k=1}^{\infty} E_k\right) = \sum_{k=1}^{\infty} P(E_k)$ (*Axiom der σ-Additivität* von P). Man kann die Ereignisse einer Wahrscheinlichkeitsalgebra eineindeutig den Elementen eines Systems B von Teilmengen einer Menge Ω derart zuordnen, daß zu den atomaren Ereignissen aus $\mathfrak{E}$ Teilmengen von Ω gehören, die jeweils nur aus einem Element von Ω bestehen. Die Elemente von Ω heißen jeweils *elementares Ereignis*. Das unmögliche Ereignis 0 entspricht der leeren Menge im System B und das sichere Ereignis I der Menge Ω. Hierbei geht $P = P(E)$ in einen Operator P über, dessen Definitionsbereich das System B von Teilmengen der Menge Ω ist. P heißt dann *Wahrscheinlichkeitsmaß* und $[\Omega, B, P]$ *Wahrscheinlichkeitsraum*.

13.2. Lineare Algebra

13.2.1. Matrizen

Ein Schema $(a_{kl})_{\substack{k=1,\ldots,m\\ l=1,\ldots,n}} = (a_{kl})_{m,n}$ aus m *Zeilen* und n *Spalten* heißt *Matrix vom Typ* (m, n),

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{pmatrix}$$

Ihre Elemente sind meist reelle oder komplexe Zahlen; in Ausnahmefällen können einige ihrer Elemente zum Beispiel auch Vektoren oder (Unter-)Matrizen sein.

13.2.1.1. Verknüpfungen. Matrizenaddition. Matrizen mit übereinstimmendem Typ können addiert werden: $A + B = (a_{kl})_{m,n} + (b_{kl})_{m,n} = (a_{kl} + b_{kl})_{m,n}$.

Multiplikation mit einer Zahl. Eine Matrix wird mit einer Zahl multipliziert, indem man jedes ihrer Elemente mit dieser Zahl multipliziert.

Matrizenmultiplikation. Hat der erste Faktor des Matrizenproduktes den Typ (m, n), der zweite Faktor den Typ (p, q), so kann die Multiplikation ausgeführt werden, falls $n = p$ ist:

$$\boldsymbol{A}\boldsymbol{B} = (a_{kl})_{m,n}(b_{rs})_{n,q} = \left(\sum_{l=1}^{n} a_{kl} b_{ls}\right)_{m,q}$$

Man sagt hierzu: In der Ergebnismatrix berechnet man dasjenige Element, das in der k-ten Zeile ($k = 1, \ldots, m$) und in der s-ten Spalte ($s = 1, \ldots, q$) steht, indem man die k-te Zeile des ersten Faktors mit der s-ten Spalte des zweiten Faktors multipliziert, d. h. *Zeilen mal Spalten*.

Determinante einer (2,2)-Matrix. Unter der Determinante einer Matrix vom Typ (2,2) versteht man – falls die Elemente der Matrix Zahlen sind – eine Zahl, nämlich

$$\det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Determinante einer (n, n)-Matrix ($n \geqq 3$). Durch *Entwicklung* nach einer beliebigen Zeile oder Spalte wird die Berechnung auf eine Determinante einer Matrix vom Typ $(n-1, n-1)$ zurückgeführt. Der Wert der Determinante hängt nicht von der Auswahl der Entwicklungszeile oder -spalte ab. Entwickelt man nach der z-ten Zeile ($z = 1, \ldots, n$) oder nach der s-ten Spalte ($s = 1, \ldots, n$), so ist $\det(a_{kl})_{n,n} = \sum_{l=1}^{n} a_{zl}A_{zl} = \sum_{k=1}^{n} a_{ks}A_{ks}$. Hierbei ist A_{zl} die *Unterdeterminante mit Vorzeichen* (*algebraisches Komplement*) von a_{zl}, d. h. man bildet zunächst aus der gegebenen Determinante die zum Element a_{zl} gehörige Unterdeterminante, indem man diejenige Zeile und diejenige Spalte streicht, denen a_{zl} angehört. Das Ergebnis wird danach mit dem Vorzeichenfaktor $(-1)^{z+l}$ multipliziert. Es ergibt sich A_{zl}.

Einheitsmatrix. Sie wird durch $E = (\delta_{kl})_{n,n}$ definiert. Das *Kronecker-Symbol* δ_{kl} hat den Wert 1 für $k = l$ und 0 für $k \neq l$. Im Bsp. 13.1.1–5 ist die Einheitsmatrix gleich dem Einselement der Gruppe.

Inverse (reziproke) Matrix A^{-1}. Im Bsp. 13.1.1–5 ist das inverse Element jeweils gleich der inversen Matrix. Genau dann ist es möglich, von der quadratischen Matrix $A = (a_{kl})_{n,n}$ die inverse Matrix A^{-1} zu bilden, wenn $\det A \neq 0$ ist. Zur Berechnung von A^{-1} kann man zunächst die Matrix $(A_{kl})_{n-1,n-1}$ herstellen, wobei A_{kl} das algebraische Komplement von a_{kl} ist. Danach vertauscht man in der Matrix (A_{kl}) die Zeilen mit den Spalten, d. h., die k-te Zeile von (A_{kl}) wird in der neuen Matrix die k-te Spalte ($k = 1, \ldots, n$). Schließlich wird die Ergebnismatrix mit $(\det A)^{-1}$ multipliziert, und es entsteht A^{-1}. In komplizierteren Fällen ist zur Berechnung von A^{-1} das Austauschverfahren zu empfehlen (vgl. 13.2.2.2.).

Transponierte Matrix A^{T}. Sie entsteht aus der gegebenen Matrix $A = (a_{kl})_{\substack{k=1,\ldots,m\\ l=1,\ldots,n}}$ durch Vertauschen von Zeilen und Spalten: $A^{\mathrm{T}} = (a_{lk})_{\substack{l=1,\ldots,n\\ k=1,\ldots,m}}$. Es gilt $(A + B)^{\mathrm{T}} = A^{\mathrm{T}} + B^{\mathrm{T}}$, $(AB)^{\mathrm{T}} = B^{\mathrm{T}}A^{\mathrm{T}}$, $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$.

Konjugiert komplexe Matrix A^*. Sie wird aus der gegebenen Matrix A gebildet, indem jedes ihrer Elemente durch das konjugiert-komplexe Element ersetzt wird. Es gilt $(A + B)^* = A^* + B^*$, $(AB)^* = A^*B^*$, $(A^*)^{\mathrm{T}} = (A^{\mathrm{T}})^*$, $(A^{-1})^* = (A^*)^{-1}$.

Assoziativ- und Distributivgesetze gelten, wenn die jeweiligen Matrizenprodukte gebildet werden können: $(AB)C = A(BC)$, $A(B + C) = AB + AC$, $(B + C)A = BA + CA$.

13.2.1.2. Matrizen mit speziellen Eigenschaften. Die folgenden Erklärungen stehen im Einklang mit 13.1.5.4.

Reelle symmetrische Matrix A. Sie wird durch $A = A^*$, $A = A^{\mathrm{T}}$ definiert.

Hermitesche Matrix A. Sie ist durch $A = (A^*)^{\mathrm{T}}$ beschrieben.

Reelle orthogonale Matrix A. Sie ist durch $A = A^*$, $AA^{\mathrm{T}} = E$, d. h. $A = A^*$, $A^{\mathrm{T}} = A^{-1}$ festgelegt. Das bedeutet, daß ihre (reellen) Spalten ein *Orthonormalsystem* bilden, d. h. es ist $a_k^{\mathrm{T}} a_l = \delta_{kl}$, falls a_k die k-te Spalte von A ist. Gleichwertig hierzu ist: Die Zeilen von A bilden ein Orthonormalsystem. Die Determinante von A ist entweder gleich 1 oder gleich -1.

Drehmatrix. Sie ist eine reelle orthogonale Matrix vom Typ (2,2) oder (3,3), deren Determinante gleich 1 ist.
(Komplexe) Unitäre Matrix $\boldsymbol{A}$. Sie ist durch $\boldsymbol{A}(\boldsymbol{A}^*)^{\mathrm{T}} = \boldsymbol{E}$, d. h. $(\boldsymbol{A}^*)^{\mathrm{T}} = \boldsymbol{A}^{-1}$ definiert. Das bedeutet, daß ihre Spalten ein Orthonormalsystem bilden, es ist also $(\boldsymbol{a}_k^*)^{\mathrm{T}}\boldsymbol{a}_l = \delta_{kl}$, falls $\boldsymbol{a}_k$ die k-te Spalte von $\boldsymbol{A}$ ist. Gleichwertig hierzu ist: Die Zeilen von $\boldsymbol{A}$ bilden ein Orthonormalsystem. Die Determinante von $\boldsymbol{A}$ hat den absoluten Betrag 1, d. h. sie ist gleich $e^{i\varphi}$ (φ ist reell).

13.2.2. Austauschverfahren

Gegeben sind die Gleichungen $y_k = \sum_{l=1}^{n} a_{kl}x_l$ $(k = 1, \ldots, m)$, kurz $\boldsymbol{y} = \boldsymbol{Ax}$ (1). Sie werden in Abb. 13.2.2–1 wiedergegeben. Gesucht ist ein gleichwertiges Gleichungssystem (2), d. h. aus (1) folgt (2) und aus (2) folgt (1), wobei x_τ und y_δ ausgetauscht sind. Es zeigt sich, daß dies im Fall $a_{\sigma\tau} \neq 0$ möglich ist. Das Ergebnis ist in Abb. 13.2.2.–2 angegeben. Die Zahlen α_{kl} sind zu bestimmen. Durch das *Austauschverfahren* (*AV*) kann diese Aufgabe schrittweise gelöst werden.
13.2.2.1. Durchführung des Austauschverfahrens. Zunächst werden in Abb. 13.2.2–1 die Elemente der Probespalte bestimmt: $a_{k,n+1} = 1 - \sum_{l=1}^{n} a_{kl}$ $(k = 1, \ldots, m)$. Mit dem *Pivot-Element* $p = a_{\sigma\tau} \neq 0$ berechnet man der Reihe nach: 1. $\alpha_{\sigma\tau} = p^{-1}$, 2. *Kellerzeile* K = *neue Pivot-Zeile*: $\alpha_{\sigma l} = -p^{-1}a_{\sigma l}$ $(l = 1, \ldots, n+1$, jedoch $l \neq \tau)$, 3. *neue Pivot-Spalte*: $\alpha_{k\tau} = p^{-1}a_{k\tau}$ $(k = 1, \ldots, m,\ k \neq \sigma)$, 4. *Rechteckregel* (Abb. 13.2.2.–1 und 13.2.2–2), $\alpha_{kl} = a_{kl} + a_{k\tau}\alpha_{\sigma l}$ $(k = 1, \ldots, m;$ $l = 1, \ldots, n+1$, jedoch $k \neq \sigma,\ l \neq \tau)$. *Probe*: im neuen Schema müssen die Zeilensummen einschließlich der Elemente der Probespalte stets gleich 1 sein. Dies ist in Abb. 13.2.2–2 rechts durch „= 1?" angedeutet.
13.2.2.2. Berechnung inverser Matrizen mit dem Austauschverfahren. Ist $m = n$ und $\det \boldsymbol{A} \neq 0$, so existiert $\boldsymbol{A}^{-1}$ und Multiplikation von (1) von links mit $\boldsymbol{A}^{-1}$ ergibt $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{y}$. Man kann also $\boldsymbol{A}^{-1}$ berechnen, indem man in (1) mit dem Austauschverfahren nacheinander alle x_l gegen alle y_k austauscht.

13.2.3. Lineare Gleichungssysteme

$\sum_{l=1}^{n} a_{kl}x_l = b_k$ $(k = 1, \ldots, m)$ (1) ist ein lineares Gleichungssystem von m Gleichungen für die n Unbekannten $x_1, \ldots, x_n$. Sind nicht alle b_k gleich null, so heißt es *inhomogen*. Das zugehörige lineare *homogene* Gleichungssystem entsteht aus (1), indem dort alle b_k durch 0 ersetzt werden.
13.2.3.1. Lösungsstruktur. Lösung. Sie ist ein Zahlen-n-Tupel $(x_1, \ldots, x_n)$, das (1) aus 13.2.3. erfüllt. Sämtliche Lösungen werden in der *allgemeinen Lösung* angegeben. Diese ist gleich einer *partikulären* (d. h. speziellen) *Lösung* des gegebenen *inhomogenen Systems* plus der *allgemeinen Lösung* des zugehörigen *homogenen Systems*. Sämtliche Lösungen des zugehörigen homogenen Systems bilden einen Vektorraum, dessen Dimension höchstens n ist.
13.2.3.2. Cramersche Regel. $\sum_{l=1}^{n} a_{kl}x_l = b_k$ $(k = 1, \ldots, m)$ lautet in Matrizenschreibweise $\boldsymbol{Ax} = \boldsymbol{b}$. Ist $m = n$ und $\det \boldsymbol{A} \neq 0$, so folgt $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, d. h., $x_l = D^{-1}D_l$ $(l = 1, \ldots, n;$ $D = \det \boldsymbol{A})$. Hierbei entsteht die Zählerdeterminante D_l aus der Nennerdeterminante D, indem die l-te Spalte von D durch die rechten Seiten des Gleichungssystems ersetzt wird. Zur numerischen Auswertung ist diese *Cramersche Regel* in den Fällen $n = m = 2$ und $n = m = 3$ zu empfehlen. Sonst wende man etwa das Austauschverfahren an (vgl. 13.2.3.4.).
13.2.3.3. Zur Lösungsherstellung des homogenen Systems im Fall $m = n$. Jetzt wird also $m = n$ und $b_k = 0$ $(k = 1, \ldots, n)$ vorausgesetzt. Ist $\det \boldsymbol{A} \neq 0$, so ergibt sich nur die *triviale Lösung* $x_l = 0$ $(l = 1, \ldots, n$; vgl. 13.2.3.2.). Ist jedoch $\det \boldsymbol{A} = 0$, so kann die allgemeine Lösung in der Gestalt $x_l = tA_{kl}$ (t ist beliebig, k ist fest) angegeben werden, falls es in $\boldsymbol{A}$ mindestens eine Zeile k gibt, für die nicht alle algebraischen Komplemente A_{kl} gleich 0 sind. Andernfalls wende man das Austauschverfahren an.
13.2.3.4. Lösungsherstellung mittels Austauschverfahren. $\sum_{l=1}^{n} a_{kl}x_l = b_k$ $(k = 1, \ldots, m)$ ist identisch mit $y_k = \sum_{l=1}^{n} a_{kl}x_l + (-b_k)\cdot 1$, wobei $y_k = 0$ $(k = 1, \ldots, m)$ ist. Das zugehörige Schema in Abb. 13.2.3–1 unterscheidet sich von

	x_1	x_τ	x_n	Pr
y_1	a_{11}	$a_{1\tau}$	a_{1n}	$a_{1,n+1}$
y_σ	$a_{\sigma 1}$	$a_{\sigma\tau}$	$a_{\sigma n}$	$a_{\sigma,n+1}$
y_m	a_{m1}	$a_{m\tau}$	a_{mn}	$a_{m,n+1}$
K	$\alpha_{\sigma 1}$	*	$\alpha_{\sigma n}$	$\alpha_{\sigma,n+1}$

Abb. 13.2.2–1 Schema der Gleichungen (1) mit *Kellerzeile* K und *Probespalte* Pr. Die Rechteckregel ist symbolisch angedeutet

–	x_1	y_σ	x_n	Pr
y_1	α_{11}	$\alpha_{1\tau}$	α_{1n}	$\alpha_{1,n+1}$ = 1?
x_τ	$\alpha_{\sigma 1}$	$\alpha_{\sigma\tau}$	$\alpha_{\sigma n}$	$\alpha_{\sigma,n+1}$ = 1?
y_m	α_{m1}	$\alpha_{m\tau}$	α_{mn}	$\alpha_{m,n+1}$ = 1?

Abb. 13.2.2–2 Schema der Gleichungen (1) nach Austausch von x_τ und y_σ. Das Ergebnis der Rechteckregel ist symbolisch angedeutet

	x_1	x_τ	x_n	1	Pr
y_1	a_{11}	$a_{1\tau}$	a_{1n}	$-b_1$	$a_{1,n+1}$
y_σ	$a_{\sigma 1}$	$a_{\sigma\tau}$	a_σ	$-b_\sigma$	$a_{\sigma,n+1}$
y_m	a_{m1}	$a_{m\tau}$	a_{mn}	$-b_m$	$a_{m,n+1}$
K	$\alpha_{\sigma 1}$	*	$\alpha_{\sigma n}$	$\alpha_{\sigma 0}$	$\alpha_{\sigma,n+1}$

Abb. 13.2.3–1 Gleichungen $y_k = \sum_{l=1}^{n} a_{kl}x_l$ mit Zusatzspalte $-b_k$, Kellerzeile K und Probespalte Pr

Abb. 13.2.2–1 nur durch die Zusatzspalte der $-b_k$. Man tausche in Abb. 13.2.3–1 möglichst viele der y_k gegen die x_l gemäß 13.2.2.1. aus. Da hier $y_k = 0$ ist, interessieren nach dem Austausch die Werte in derjenigen Spalte nicht, die oben durch das ausgetauschte y_k gekennzeichnet ist. Folglich ist es ökonomisch, diese Spalte wegzulassen und α_{kr} $(k = 1, \ldots, m)$ erst gar nicht zu berechnen. Man bezeichnet diese Methode als *Austauschverfahren mit Spaltentilgung* (AVS). Zur Probe setzt man $a_{k,n+1} = -\left\{\sum_{l=1}^{n} a_{kl} + (-b_k)\right\}$ und prüft im neuen Schema, ob die Zeilensumme einschließlich der Elemente aus der Probespalte gleich 1 oder gleich 0 ist, je nachdem am Kopf der Zeile ein x oder y steht.

1. Fall: Alle $y_1, \ldots, y_m$ können ausgetauscht werden. Sind dann in der Kopfzeile des letzten Schemas auch keine x-Größen da (die y-Größen wurden ja getilgt), so liest man genau eine Lösung $(x_1, \ldots, x_n)$ ab. Sind dagegen in der Kopfzeile des letzten Schemas x-Größen vorhanden – ihre Anzahl betrage r – so wird eine jede von diesen gleich einer willkürlichen Zahl t_ϱ $(\varrho = 1, \ldots, r)$ gesetzt. Die übrigen x-Werte können dann aus dem letzten Schema entnommen werden. Der Vektorraum der Lösungen des zugehörigen homogenen Systems ist jetzt r-dimensional und ist eine Teilmenge von Bsp. 13.1.4–2.

2. Fall: Es können nicht alle $y_1, \ldots, y_m$ ausgetauscht werden. Dann ist das Gleichungssystem nur lösbar, wenn im letzten Schema in der Zeile des jeweils nicht austauschbaren y nicht nur alles Nullen stehen, die zu x-Spalten gehören (es sind dort Nullen, sonst könnte ja doch weiter ausgetauscht werden), sondern auch jeweils eine Null steht, die zur 1-Spalte gehört. Die Lösungen entnimmt man dann wie im ersten Fall.

Austauschverfahren mit Spalten- und Zeilentilgung (AVSZ). Hierzu berechnet man die Kellerzeile und arbeitet beim Herstellen des neuen Schemas nur mit der Rechteckregel. Aus dem letzten Schema entnimmt man wie bisher entweder die Nichtlösbarkeit oder im Fall der Lösbarkeit diejenigen x-Werte (sie stehen in der Kopfzeile), die gleich beliebigen Zahlen t_ϱ zu setzen sind. Die übrigen x-Werte ergeben sich nacheinander aus der Gesamtheit der Kellerzeilen und damit der jeweils neuen Pivotzeilen.

13.2.4. Eigenwertaufgaben bei Matrizen

Gesucht sind komplexe Zahlen λ derart, daß $\boldsymbol{A u} = \lambda \boldsymbol{u}$, $\boldsymbol{A}$ Typ (n, n); $\boldsymbol{u}$ Typ $(n, 1)$, *nichttriviale* Lösungen $\boldsymbol{u}$ besitzt, d. h. Lösungen $\boldsymbol{u}$ deren Elemente nicht alle gleich 0 sind. Diese heißen Eigenwerte, die dazugehörigen Lösungen $\boldsymbol{u}$ *Eigenvektoren* (es sind dies hier einspaltige Matrizen).

13.2.4.1. Eigenwertbestimmung. $\boldsymbol{Au} = \lambda \boldsymbol{u}$, d. h. $(\boldsymbol{A} - \lambda \boldsymbol{E})\boldsymbol{u} = \boldsymbol{0}$ (1), [$\boldsymbol{0}$ Typ $(n, 1)$, alle Elemente sind gleich 0] hat genau dann Lösungen $\boldsymbol{u} \neq \boldsymbol{0}$, wenn λ die *Eigenwertgleichung* $\det(\boldsymbol{A} - \lambda \boldsymbol{E}) = \boldsymbol{0}$ (2) erfüllt. Auf der linken Seite von (2) steht ein *Polynom* $P_n(\lambda)$ vom n-ten Grad: $P_n(\lambda) = a_n\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$ (hier ist $a_n = (-1)^n$). Hat man eine Lösung $\lambda = \lambda_1$ von (2), also einen Eigenwert λ_1 gefunden, z. B. durch Raten, so ergibt die Division von $P_n(\lambda)$ durch $\lambda - \lambda_1$ ein Polynom $b_{n-1}\lambda^{n-1} + \ldots + b_1\lambda + b_0$, dessen Nullstellen die weiteren Lösungen von (2) sind. Die obige Division geschieht zweckmäßig mit dem *Horner-Schema* (Abb. 13.2.4–1). Die untereinanderstehenden Zahlen in der Abb. sind zu addieren: $b_{n-1} = a_n$, $b_{n-2} = a_{n-1} + \lambda_1 b_{n-1}$, …, $b_0 = a_1 + \lambda_1 b_1$, $a_0 + \lambda_1 b_0 = 0$. Sind $\lambda_1, \ldots, \lambda_r$ die voneinander

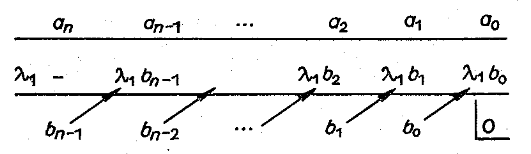


Abb. 13.2.4–1 Horner-Schema, die Pfeile deuten die Multiplikation mit λ_1 an

verschiedenen Lösungen von (2), so gilt $P_n(\lambda) = a_n(\lambda - \lambda_1)^{l_1}(\lambda - \lambda_2)^{l_2} \ldots (\lambda - \lambda_r)^{l_r}$. Hierbei heißt l_ϱ die *Vielfachheit* von λ_ϱ $(\varrho = 1, \ldots, r)$. Nach dem *Fundamentalsatz der Algebra* gilt $l_1 + l_2 + \ldots + l_r = n$.

13.2.4.2. Eigenräume. Durch Einsetzen des Eigenwertes $\lambda = \lambda_\varrho$ $(\varrho = 1, \ldots, r)$ in (1) aus 13.2.4.1. ergibt sich ein lineares homogenes Gleichungssystem für die Elemente der zu λ_ϱ gehörigen $\boldsymbol{u} = \boldsymbol{u}_\varrho$. Die Lösungen $\boldsymbol{u}_\varrho$ bilden einen Vektorraum, den zu λ_ϱ gehörigen *Eigenraum*. Für seine Dimension d gilt $1 \leqq d \leqq l_\varrho$.

Eigenraum von Eigenwertaufgaben hermitescher Matrizen. Die Eigenwerte aus Eigenwertaufgaben hermitescher Matrizen sind alle reell.

Die Dimension des zu λ_ϱ gehörigen Eigenraumes ist *gleich* der Vielfachheit l_ϱ von λ_ϱ. Eigenräume (Basen $\boldsymbol{u}_k$ bzw. $\boldsymbol{v}_l$), die zu verschiedenen Eigenwerten λ_ϱ, $\lambda_{\varrho'}$, $(\varrho \neq \varrho')$ gehören, sind zueinander orthogonal, d. h. es gilt

$(\boldsymbol{u}_k^*)^T \boldsymbol{v}_l = 0$; $[k = 1, \ldots, l_\varrho;\ l = 1, \ldots, l_{\varrho'}]$.

13.2.5. Vektoralgebra

13.2.5.1. Verknüpfungen. Gebundene Vektoren sind gegeben durch Angriffspunkt, Richtung einschließlich Orientierung und dem absoluten Betrag $|\boldsymbol{a}|$. Sie können durch Pfeile dargestellt werden, die im Angriffspunkt angeheftet sind, in die gegebene Richtung weisen und deren Längen $|\boldsymbol{a}|$ entsprechen. Der *Nullvektor* $\boldsymbol{0}$ hat den absoluten Betrag 0 ($|\boldsymbol{0}| = 0$), als seine Richtung kann jede Richtung angesehen werden.

Addition. Sie geschieht gemäß der bekannten Kräfteparallelogrammkonstruktion. Alle in einem Punkt gebundenen Vektoren bilden mit der Addition als Verknüpfung eine Abelsche Gruppe (vgl. 13.1.1.5.).

Multiplikation mit einer reellen Zahl. Ist $c > 0$ bzw. $c < 0$, so ist $c\boldsymbol{a}$ gleich- bzw. entgegengesetzt gerichtet zu $\boldsymbol{a}$ und es ist $|c\boldsymbol{a}| = |c|\,|\boldsymbol{a}|$.

Skalarprodukt. $\boldsymbol{a} \cdot \boldsymbol{b} = |\boldsymbol{a}|\,|\boldsymbol{b}| \cos \varphi$ (vgl. Bsp. 13.1.4–6 und 13.1.4–7).

Vektorprodukt (Kreuzprodukt) $\boldsymbol{a} \times \boldsymbol{b}$. Der Vektor $\boldsymbol{a} \times \boldsymbol{b}$ steht sowohl auf $\boldsymbol{a}$ als auch auf $\boldsymbol{b}$ senkrecht. Dreht man $\boldsymbol{a}$ auf dem kürzesten Weg in die Richtung von $\boldsymbol{b}$ und schreitet man hierbei in Richtung von $\boldsymbol{a} \times \boldsymbol{b}$ fort, so ergibt sich eine *Rechtsschraube*. Es ist $|\boldsymbol{a} \times \boldsymbol{b}| = |\boldsymbol{a}|\,|\boldsymbol{b}| \sin \varphi$, wobei φ $(0 \leqq \varphi \leqq \pi)$ den Winkel zwischen $\boldsymbol{a}$ und $\boldsymbol{b}$ bedeutet.

Zerlegungssatz. $\boldsymbol{a} \times (\boldsymbol{b} \times \boldsymbol{c}) = \boldsymbol{b}(\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{c}(\boldsymbol{a} \cdot \boldsymbol{b})$, $(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} = \boldsymbol{b}(\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{a}(\boldsymbol{b} \cdot \boldsymbol{c})$ (vgl. 13.1.2.2. mit Bsp. 13.1.2–1).

Spatprodukt. $(\boldsymbol{abc}) = \boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c}) = (\boldsymbol{a} \times \boldsymbol{b}) \cdot \boldsymbol{c}$ ist gleich dem vorzeichenbehafteten Rauminhalt des von den (im gleichen Punkt angehefteten) Vektoren $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ aufgespannten Spats. Er ist >0 bzw. <0 bzw. $=0$, je nachdem, ob $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ (in dieser Reihenfolge) ein Rechts- bzw. Linkssystem bilden bzw. in einer Ebene liegen. Ein *Rechts-* bzw. *Linkssystem* $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ bedeutet, daß Daumen, Zeigefinger, Mittelfinger der ge-

spreizten *rechten* bzw. *linken* Hand in die Richtungen von $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$ zeigen.

Freie Vektoren $\boldsymbol{a}$. Gebundene Vektoren, die sich nur im Angriffspunkt unterscheiden, stellen ein und denselben *freien Vektor* dar. Beim Ausführen der Rechenoperationen bei freien Vektoren kann man sie durch Parallelverschiebung an einem (beliebigen) gemeinsamen Angriffspunkt anheften.

Linienflüchtige Vektoren dürfen – ohne sie zu ändern – von einem ursprünglichen Angriffspunkt aus nur längs einer Linie verschoben werden, die einerseits durch diesen Angriffspunkt geht und andererseits die Richtung des Vektors besitzt. Die Addition linienflüchtiger Vektoren ergibt nicht immer wieder einen linienflüchtigen Vektor (Kräftepaar)!

Ortsvektoren sind gebundene Vektoren, deren Angriffspunkt der Koordinatenursprung ist. Dagegen sind Differenzen von Ortsvektoren freie Vektoren.

13.2.5.2. Vektoralgebra in kartesischen Koordinaten. Basisvektoren $\boldsymbol{e}_k$ ($k = 1, 2, 3$) (vgl. Bsp. 13.1.4–7) sind keine echten Vektoren, denn sie hängen von der Wahl des vorliegenden kartesischen Koordinatensystems (x_1, x_2, x_3) ab.

Zerlegung eines Vektors $\boldsymbol{a}$ relativ zur orthonormalen Basis $\boldsymbol{e}_k$. (Vgl. Fourier-Entwicklung aus 13.1.4.1.) $\boldsymbol{a} = \sum_{k=1}^{3} a_k \boldsymbol{e}_k$ mit den *Koordinaten a_k des Vektors $\boldsymbol{a}$*: $a_k = \langle \boldsymbol{e}_k | \boldsymbol{a} \rangle = \boldsymbol{e}_k \cdot \boldsymbol{a}$ (Abb. 13.2.5–1). Anstatt $\boldsymbol{a} = \sum_{k=1}^{3} a_k \boldsymbol{e}_k$ schreibt man kurz $\boldsymbol{a} = a_k \boldsymbol{e}_k$ mit der *Einsteinschen Summationsverabredung*: Treten in einem Produkt zwei einander gleiche Indizes auf, so ist über diese Indizes zu summieren (hier von 1 bis 3). Treten jedoch in einem Produkt mehr als zwei einander gleiche Indizes auf, so ist die Verabredung nicht anwendbar.

Translation des kartesischen Koordinatensystems. Bei dieser Parallelverschiebung ändern sich weder die Basisvektoren $\boldsymbol{e}_k$ noch die Koordinaten a_k von Vektoren $\boldsymbol{a}$. Der Ortsvektor $\boldsymbol{r}$ geht bei der Translation um $\boldsymbol{r}_0$ in $\boldsymbol{r} - \boldsymbol{r}_0$ über.

Drehung des kartesischen Koordinatensystems in ein neues System ($\tilde{x}_1$, $\tilde{x}_2$, $\tilde{x}_3$), Abb. 13.2.5–2. Die neuen Basisvektoren $\tilde{\boldsymbol{e}}_k$ gehen aus den alten Basisvektoren $\boldsymbol{e}_k$ gemäß $\tilde{\boldsymbol{e}}_k = D_{kl} \boldsymbol{e}_l$ ($k = 1, 2, 3$; über l summieren!) hervor, wobei (D_{kl}) eine Drehmatrix ist (vgl. 13.2.1.2.). Hierbei ist D_{kl} gleich dem Kosinus des Winkels zwischen der $\tilde{x}_k$-Achse und der x_l-Achse. Die alten bzw. neuen Koordinaten a_k bzw. $\tilde{a}_k$ eines echten (d. h. vom Koordinatensystem unabhängigen) Vektors $\boldsymbol{a} = a_k \boldsymbol{e}_k = \tilde{a}_k \tilde{\boldsymbol{e}}_k$ hängen wie $\boldsymbol{e}_k$ und $\tilde{\boldsymbol{e}}_k$ zusammen, d. h. es gilt $\tilde{a}_k = D_{kl} a_l$.

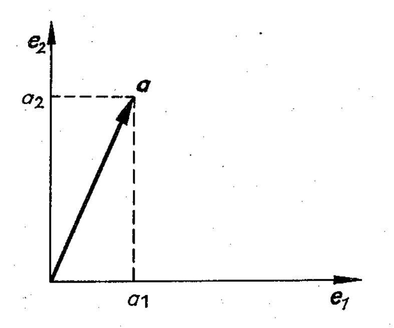


Abb. 13.2.5–1 Zerlegung eines Vektors, dargestellt in der Ebene

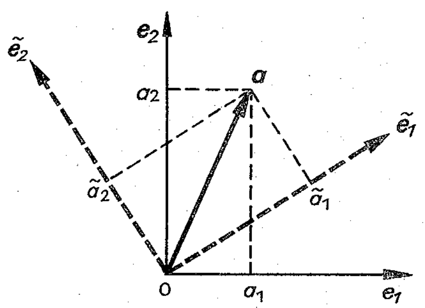


Abb. 13.2.5–2 Drehung des Koordinatensystems um die im Punkt 0 auf der Zeichenebene senkrecht stehende Achse

Skalarprodukt. Die Zahl $\boldsymbol{a} \cdot \boldsymbol{b} = a_k \boldsymbol{e}_k \cdot b_l \boldsymbol{e}_l = a_k b_l \boldsymbol{e}_k \cdot \boldsymbol{e}_l = a_k b_l \delta_{kl} = a_k b_k = \tilde{a}_k \tilde{b}_k$ ist *unabhängig* von der Wahl des kartesischen Koordinatensystems.

Vektorprodukt. Im kartesischen Rechtssystem (d.h. die x_1-, x_2-, x_3-Achse bilden ein Rechtssystem, vgl. 13.2.5.1.) gilt

$$\boldsymbol{a} \times \boldsymbol{b} = \begin{vmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \boldsymbol{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix} = \varepsilon_{klm} a_k b_l \boldsymbol{e}_m .$$

ε_{klm} ist das *Levi-Civita-Symbol*, auch *Epsilon-Tensor* genannt. $\varepsilon_{klm} = 1$ für $(k, l, m) = (1, 2, 3)$ oder $(2, 3, 1)$ oder $(3, 1, 2)$ (zyklische Vertauschung!); $\varepsilon_{klm} = -1$ für $(k, l, m) = (1, 3, 2)$ oder $(3, 2, 1)$ oder $(2, 1, 3)$; $\varepsilon_{klm} = 0$, falls mindestens zwei der Indizes einander gleich sind.

Spatprodukt. $$(\boldsymbol{a}\boldsymbol{b}\boldsymbol{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = \varepsilon_{klm} a_k b_l c_m .$$

Zerlegungssatz. Dieser ist äquivalent zu $\varepsilon_{klm} \varepsilon_{pqm} = \delta_{kp} \delta_{lq} - \delta_{kq} \delta_{lp}$.

13.2.5.3. Vektoralgebra in orthogonalen krummlinigen Koordinaten. Krummlinige Koordinaten. Bei vorliegendem kartesischen Rechtssystem (x_1, x_2, x_3) werden durch die drei reellen Funktionen von jeweils drei reellen unabhängigen Variablen $x_k = x_k(u_1, u_2, u_3)$ (1) ($k = 1, 2, 3$) *krummlinige Koordinaten* (u_1, u_2, u_3) eingeführt. Läßt man nur u_l (l = fest) variieren, während die beiden anderen krummlinigen Koordinaten festgehalten werden, wird durch (1) eine Parameterdarstellung der *u_l-Koordinatenlinie* geliefert. Die partiellen Ableitungen $\partial x_k / \partial u_l$ (l = fest, $k = 1, 2, 3$) ergeben im (x_1, x_2, x_3)-System die Koordinaten eines Tangentenvektors $\boldsymbol{g}_{u_l}$ an die u_l-Linie. Der zugehörige Einheitsvektor lautet $\boldsymbol{e}_{u_l} = |\boldsymbol{g}_{u_l}|^{-1} \boldsymbol{g}_{u_l}$.

Orthogonale krummlinige Koordinaten liegen vor, wenn die Tangenteneinheitsvektoren $\boldsymbol{e}_{u_l}$ an die u_l-Koordinatenlinien ein Orthonormalsystem bilden: $\boldsymbol{e}_{u_l} \cdot \boldsymbol{e}_{u_k} = \delta_{kl}$.

Bsp. 13.2.5–1. *Zylinderkoordinaten* $(u_1, u_2, u_3) = (r_\perp, \varphi, z)$: $x_1 = r_\perp \cos\varphi$, $x_2 = r_\perp \sin\varphi$, $x_3 = z$. Hierbei ist $r_\perp$ der (senkrechte) Abstand des Punktes (x_1, x_2, x_3) von der x_3-Achse. Der Winkel φ wird in der Ebene x_3 = const. gemessen, und zwar von der Richtung der x_1-Achse bis zur Richtung desjenigen Vektors, der im Punkt $(0, 0, x_3)$ angreift und zum Punkt (x_1, x_2, x_3) weist. Es ist $\boldsymbol{e}_{u_1} = \boldsymbol{e}_{r_\perp} = \cos\varphi\, \boldsymbol{e}_1 + \sin\varphi\, \boldsymbol{e}_2$, $\boldsymbol{e}_{u_2} = \boldsymbol{e}_\varphi = -\sin\varphi\, \boldsymbol{e}_1 + \cos\varphi\, \boldsymbol{e}_2$, $\boldsymbol{e}_{u_3} = \boldsymbol{e}_z = \boldsymbol{e}_3$. Zylinderkoordinaten sind also orthogonal.

Bsp. 13.2.5–2. *Kugelkoordinaten* $(u_1, u_2, u_3) = (r, \vartheta, \varphi)$: $x_1 = r \sin\vartheta \cos\varphi$, $x_2 = r \sin\vartheta \sin\varphi$, $x_3 = r \cos\vartheta$. Hierbei ist r

der Abstand des Punktes (x_1, x_2, x_3) vom Koordinatenursprung. Der Winkel ϑ wird zwischen der x_3-Achse und dem Ortsvektor $x_k e_k$ gemessen. Der Winkel φ ist so wie im Bsp. 13.2.5–1 definiert. Es ist

$$e_{u_1} = e_r = \sin\vartheta\cos\varphi\, e_1 + \sin\vartheta\sin\varphi\, e_2 + \cos\vartheta\, e_3,$$
$$e_{u_2} = e_\vartheta = \cos\vartheta\cos\varphi\, e_1 + \cos\vartheta\sin\varphi\, e_2 - \sin\vartheta\, e_3,$$
$$e_{u_3} = e_\varphi = -\sin\varphi\, e_1 + \cos\varphi\, e_2.$$

Kugelkoordinaten sind also orthogonal.

Bsp. 13.2.5–3. Werden krummlinige Koordinaten (u_1, u_2, u_3) durch $x_1 = u_1$, $x_2 = u_1 + u_2$, $x_3 = u_1 + u_2 + u_3$ eingeführt, so sind die u_l-Koordinatenlinien ($l = 1, 2, 3$) stets Geraden. Es liegt aber kein orthogonales System (u_1, u_2, u_3) vor.

Vektorfelder. Wird jedem Punkt P eines gewissen Bereiches im Raum ein gebundener Vektor zugeordnet, dessen Angriffspunkt P ist, so spricht man von einem *Vektorfeld* $a = a(P)$. Liegt ein orthogonales krummliniges Koordinatensystem (u_1, u_2, u_3) vor, so wird jeder Punkt des obigen Bereiches durch ein Wertetripel (u_1, u_2, u_3) gegeben. Das im Punkt P angeheftete Orthonormalsystem e_{u_l} ($l = 1, 2, 3$) übernimmt die Rolle von e_k ($k = 1, 2, 3$) beim Aufbau einer Vektoralgebra gemäß 13.2.5.2. für Vektoren a, die im Punkt P angeheftet sind.

13.2.6. Tensoralgebra

13.2.6.1. Definition von Tensoren im Euklidischen Raum.

Tensoren nullter Stufe sind skalare Größen, die vom Koordinatensystem unabhängig sind. So ist $a \cdot b$ (a, b = Vektoren) ein Tensor nullter Stufe (vgl. 13.2.5.2., Skalarprodukt). a_1 aus $a = a_k e_k$ ist jedoch kein Tensor nullter Stufe.

Tensoren erster Stufe sind (echte) Vektoren.

Tensoren zweiter Stufe. Es handelt sich um lineare Operatoren T (vgl. 13.1.5.), deren Definitionsbereich und Wertebereich in der Menge der Vektoren enthalten sind.

Bsp. 13.2.6–1. Eine Masse im Volumen V, die sich mit der Winkelgeschwindigkeit $|\omega|$ um eine solche Achse dreht, die durch den Ursprung eines vorliegenden kartesischen Koordinatensystems (x_1, x_2, x_3) geht und in die Richtung des Winkelgeschwindigkeitsvektors ω weist, hat den Gesamtdrehimpulsvektor L, wobei $L = T\omega$ gilt. Hierbei ist T ein Tensor zweiter Stufe, der *Trägheitstensor*. Mit der Massendichte $\varrho = \varrho\,(x_1, x_2, x_3)$ ist $T\omega$ gleich dem über V zu erstreckendem Raumintegral $\iiint \varrho\,[r \times (\omega \times r)]\,dV = \iiint \varrho\,[\omega(r \cdot r) - r(r \cdot \omega)]\,dV$ (r = Ortsvektor des variablen Integrationspunktes (x_1, x_2, x_3)).

Bsp. 13.2.6–2. Man zerschneide (im Gedankenexperiment) ein belastetes Kontinuum und ersetze den Einfluß des entstehenden einen Teilstückes auf das andere Teilstück durch innere Kräfte längs der Schnittfläche. Betrachtet man an der Stelle (x_1, x_2, x_3) ein (kleines) Flächenelement dA der Schnittfläche, so wirkt auf dieses der Kraftvektor $\sigma n\, dA$. Hierbei ist n ein auf dA senkrecht stehender, nach außen gerichteter Vektor mit $|n| = 1$. $\sigma = \sigma\,(x_1, x_2, x_3)$ ist ein Tensor zweiter Stufe, der *Spannungstensor*.

Tensoren n-ter Stufe ($n \geqq 3$) sind lineare Operatoren T. Als Definitionsbereich kann man Tensoren p-ter Stufe ($p = 1, \ldots, n-1$) und als Wertebereich Tensoren q-ter Stufe ($q = 1, \ldots, n-1$) mit $p + q = n$ wählen.

Bsp. 13.2.6–3. Mit dem Spannungstensor σ, also einem Tensor zweiter Stufe, und dem *Verzerrungstensor* γ – das ist ebenfalls ein Tensor zweiter Stufe – kann das *Hookesche Gesetz* durch $\sigma = E\gamma$ angegeben werden, wobei der *Elastizitätstensor* E ein Tensor vierter Stufe ist.

13.2.6.2. Darstellung von Tensoren im kartesischen Rechtssystem.

Tensoren zweiter Stufe T. Es ist $w = Tu$ gleichbedeutend mit $w_k e_k = T(u_l e_l) = u_l T(e_l)$ (1). Die Darstellung des Vektors $T(e_l)$ relativ zur Basis e_k ($k = 1, 2, 3$) wird durch $T(e_l) = T_{kl} e_k$ (2) angegeben, d. h. der Vektor e_l wird durch T in einen Vektor abgebildet, dessen Koordinaten relativ zur Basis e_k ($k = 1, 2, 3$) die l-te Spalte der Matrix (T_{kl}) bilden. Aus (1) und (2) folgt $w_k = T_{kl} u_l$ oder in Matrizenschreibweise $w = Tu$ mit $T = (T_{kl})$.

Bsp. 13.2.6–4. Relativ zum kartesischen Rechtssystem wird der Trägheitstensor T aus Bsp. 13.2.6–1 durch die Matrix (T_{kl}) dargestellt mit $T_{kl} = \iiint \varrho\,(x_p x_p \delta_{kl} - x_k x_l)\,dV$, also $T_{11} = \iiint \varrho\,(x_2^2 + x_3^2)\,dV$, $T_{12} = -\iiint \varrho x_1 x_2\,dV, \ldots$.

Bsp. 13.2.6–5. Bei *festgehaltenem* kartesischen Rechtssystem werden alle Punkte (x_1, x_2, x_3) um eine durch den Koordinatenursprung gehende Achse, deren Richtung durch $d = d_k e_k$ mit $|d| = 1$ festgelegt ist, um den Winkel $\alpha > 0$ gedreht. Dabei sei d so orientiert, daß die Drehung gekoppelt mit Fortschreiten in Richtung von d eine Rechtsschraube liefert. Der Bildpunkt (y_1, y_2, y_3) von (x_1, x_2, x_3) ergibt sich durch $y_k = T_{kl} x_l$. Die Matrix (T_{kl}) ist eine Drehmatrix (vgl. 13.2.1.2.). Sie stellt einen Tensor zweiter Stufe, den *Drehtensor* (Abb. 13.2.6–1), relativ zum vorliegenden Rechtssystem (x_1, x_2, x_3) dar. Es ist $T_{kl} = d_k d_l + (\delta_{kl} - d_k d_l)\cos\alpha - \varepsilon_{klm} d_m \sin\alpha$ (3). Vektoren, die mit den Basisvektoren e_k ($k = 1, 2, 3$) zusammenfallen, gehen wegen (2) hierbei in $\tilde{e}_l = T_{kl} e_k$ ($l = 1, 2, 3$) oder – wenn man die Bezeichnungen von k und l vertauscht – in $\tilde{e}_k = T_{lk} e_l$ ($k = 1, 2, 3$) über. Führt man also mit obiger Drehung ein neues Koordinatensystem $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ ein, so ist die Drehmatrix (D_{kl}) aus 13.2.5.2. gleich der zu (T_{kl}) transponierten Matrix. Sie entsteht auch aus (3) durch Ersetzen von α durch $-\alpha$.

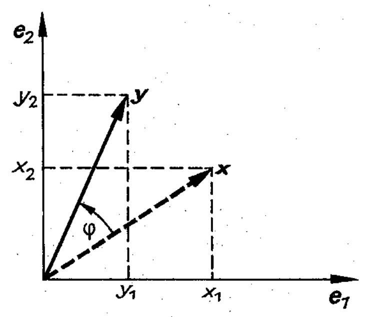


Abb. 13.2.6–1 Drehtensor, $d = e_3$ steht senkrecht auf der Zeichenebene

Translation des kartesischen Koordinatensystems. Bei dieser Parallelverschiebung ändern sich weder die Basisvektoren e_k noch die Koordinaten von Tensoren.

Drehung des kartesischen Koordinatensystems. Mit $\tilde{e}_k = D_{kl} e_l$ (vgl. 13.2.5.2.) gilt $T_{pq} = D_{pk} D_{ql} T_{kl}$, d. h. in Matrizenschreibweise $T = DTD^{\mathrm{T}}$.

Hauptachsentransformation. Ein symmetrischer Tensor T zweiter Stufe habe relativ zu einem kartesischen Koordinatensystem (x_1, x_2, x_3) die Koordinaten $T_{kl} = T_{lk}$. Dreht man das Koordinatensystem derart, daß die Basisvektoren $\tilde{e}_k$ des gedrehten Systems $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$ Eigenvektoren der Eigenwertaufgabe $Tu = \lambda u$ (vgl. 13.2.4.1.) sind, so wird der Tensor T im $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$-System durch eine Matrix $\tilde{T}_{kl}$ dargestellt, deren Elemente außerhalb der Hauptdiagonale alle gleich null sind, während in der Hauptdiagonale die Eigenwerte der Ei-

genwertaufgabe stehen – jeweils so oft, wie ihre Vielfachheit angibt.

Zerlegung eines Tensors zweiter Stufe in seinen symmetrischen und antimetrischen Anteil. Mit $T_{(kl)} = (1/2)(T_{kl} + T_{lk})$, $T_{[kl]} = (1/2)(T_{kl} - T_{lk})$ und $T_m = (1/2)\varepsilon_{mkl}T_{kl}$ gilt $T_{kl} = T_{(kl)} + T_{[kl]} = T_{(kl)} + \varepsilon_{klm}T_m$ (vgl. Bsp. 13.2.6–10).

Tensoren dritter Stufe können relativ zu einem kartesischen Rechtssystem durch T_{mkl} beschrieben und etwa durch $w_m = T_{mkl}u_{kl}$ eingeführt werden. Dabei stellt u_{kl} einen Tensor zweiter Stufe und w_m einen Tensor erster Stufe relativ zum vorliegenden Koordinatensystem dar.

Bsp. 13.2.6–6. Mit $\boldsymbol{a} = a_k\boldsymbol{e}_k$, $\boldsymbol{b} = b_l\boldsymbol{e}_l$ ist die Matrix (a_kb_l) relativ zum Koordinatensystem ein Tensor zweiter Stufe. Wegen $\boldsymbol{a} \times \boldsymbol{b} = \varepsilon_{mkl}a_kb_l\boldsymbol{e}_m$ ($\varepsilon_{mkl} = \varepsilon_{klm} = \varepsilon_{lmk}$, vgl. 13.2.5.2.) stellt ε_{mkl} relativ zum vorliegenden Rechtssystem einen Tensor dritter Stufe dar. Damit ist der Name *Epsilontensor* verständlich.

Tensoren vierter und höherer Stufe. Die Einführung wird bereits deutlich, wenn man gemäß Bsp. 13.2.6–3 $\sigma_{kl} = E_{klmp}\gamma_{mp}$ notiert.

13.2.6.3. Darstellung von Tensoren in orthogonalen, krummlinigen Koordinaten. Auf der Grundlage der Vektoralgebra in orthogonalen, krummlinigen Koordinaten (vgl. 13.2.5.3.) kann die Theorie aus 13.2.6.2. auf die Darstellung von Tensoren in orthogonalen, krummlinigen Koordinaten übertragen werden, wobei die Rolle der e_k von e_{u_k} übernommen wird.

13.2.6.4. Addition und Multiplikation von Tensoren. Addition. Wenn man relativ zu einem kartesischen Rechtssystem oder orthogonalen, krummlinigen Koordinatensystem die Koordinaten zweier Tensoren n-ter Stufe addiert, so ergeben sich die Koordinaten eines Tensors n-ter Stufe.

Multiplikation. Wenn man relativ zu einem kartesischen Rechtssystem oder orthogonalem krummlinigen Koordinatensystem die Koordinaten eines Tensors n-ter Stufe mit denen eines Tensors m-ter Stufe multipliziert, so ergeben sich die Koordinaten eines Tensors der Stufe $n + m$.

Bsp. 13.2.6–7. Sind a_k und b_l die Koordinaten von Tensoren erster Stufe, so sind $\varepsilon_{pqm}a_kb_l$ die Koordinaten eines Tensors fünfter Stufe.

13.2.6.5. Verjüngung und Überschiebung von Tensoren. Verjüngung. Werden in den Koordinaten eines Tensors n-ter Stufe ($n \geqq 2$) zwei Indizes einander gleich gesetzt und wird gemäß der Einsteinschen Summationsverabredung über diese summiert, so entstehen die Koordinaten eines Tensors $(n\text{-}2)$-ter Stufe.

Überschiebung. Das ist die Verjüngung eines Produktes zweier Tensoren, wobei ein Index des einen Faktors mit einem Index des anderen Faktors gleichgesetzt wird.

Bsp. 13.2.6–8. Zweimalige Verjüngung des Tensors aus Bsp. 13.2.6–7 führt zu den Koordinaten $\varepsilon_{klm}a_kb_l$ des Vektorproduktes $\boldsymbol{a} \times \boldsymbol{b}$.

13.2.6.6. Differentiation von Tensoren. Relativ zu einem kartesischen Rechtssystem bilden die partiellen Ableitungen der Koordinaten eines Tensorfeldes n-ter Stufe die Koordinaten eines Tensorfeldes $(n + 1)$-ter Stufe. Diese Tatsache ist nicht auf orthogonale krummlinige Koordinaten übertragbar.

Bsp. 13.2.6–9. Bildet man von den relativ zu einem kartesischen Rechtssystem gegebenen Koordinaten σ_{kl} eines Spannungstensorfeldes sämtlich partiellen Ableitungen – Schreibweisen: $\dfrac{\partial\sigma_{kl}(x_1, x_2, x_3)}{\partial x_m} = \dfrac{\partial}{\partial x_m}\sigma_{kl} = \partial_m\sigma_{kl} = \sigma_{kl,m}$ – so entstehen die Koordinaten eines Tensorfeldes dritter Stufe.

Durch die Verjüngung $\sigma_{kl,l}$ ergeben sich die Koordinaten eines Vektors, der ein wesentlicher Bestandteil der Gleichgewichtsbedingungen oder auch der Bewegungsgleichungen in der Kontinuumsmechanik ist.

13.2.6.7. Vektoranalysis in kartesischen Koordinaten.

Gradient. $\mathbf{grad}\, U(x_1, x_2, x_3) = \dfrac{\partial U}{\partial x_k}\boldsymbol{e}_k = U_{,k}\boldsymbol{e}_k$

Divergenz. $\operatorname{div} \boldsymbol{v} = \operatorname{div}[v_k(x_1, x_2, x_3)\boldsymbol{e}_k] = v_{k,k}$

Laplace-Operator. $\Delta U = \operatorname{div}(\mathbf{grad}\, U) = (U_{,k})_{,k}$

Rotation.

$\mathbf{rot}\, \boldsymbol{v} = \mathbf{rot}[v_k(x_1, x_2, x_3)\boldsymbol{e}_k] = \varepsilon_{klm}\partial_k v_l \boldsymbol{e}_m = \varepsilon_{klm}v_{l,k}\boldsymbol{e}_m$.

Bsp. 13.2.6–10. Ist $\boldsymbol{v} = v_k(x_1, x_2, x_3, t)\boldsymbol{e}_k$ (t = Zeit) das Geschwindigkeitsvektorfeld eines sich bewegenden Kontinuums, so ergibt die Zerlegung von $v_{k,l}$ in symmetrischen und antimetrischen Anteil: $v_{k,l} = v_{(k,l)} + \varepsilon_{klm}(-\omega_m)$. Hierbei sind $v_{(k,l)}$ die Koordinaten des *Deformationsgeschwindigkeitstensors* und $\omega_m = -(1/2)\varepsilon_{mkl}v_{k,l} = (1/2)\varepsilon_{lkm}v_{k,l} = (1/2)\varepsilon_{klm}v_{l,k}$ die Koordinaten des *Wirbelvektors* $\boldsymbol{\omega} = \omega_m\boldsymbol{e}_m = (1/2)\,\mathbf{rot}\, \boldsymbol{v}$.

13.3. Differentialrechnung

13.3.1. Der Funktionsbegriff

M_1 und M_2 seien (nichtleere) Mengen. $D(F)$ sei eine Teilmenge von M_1: $D(F) \subseteqq M_1$ (eine Menge wird auch als Teilmenge von sich selbst aufgefaßt: $M_1 \subseteqq M_1$). *Funktion F* (auch *Operator* oder *Transformation* oder *Abbildung*) ist eine *Zuordnungsvorschrift*, die jedem Element u aus dem *Definitionsbereich* $D(F)$ genau ein Element w aus M_2 zuordnet [Schreibweise: $w = F(u)$ oder auch $w = Fu$]. Durchläuft u alle Elemente von $D(F)$, so durchläuft $w = F(u)$ eine Teilmenge von M_2, die *Wertebereich* von F genannt wird. Die Funktion F heißt *umkehrbar* (*invertierbar*), wenn die Gleichung $F(u) = w$ für jedes w aus dem Wertebereich von F genau eine Lösung u besitzt. Man schreibt dann $u = F^{-1}(w)$, wobei F^{-1} *Umkehrfunktion* von F heißt.

13.3.1.1. Reelle Funktionen einer reellen unabhängigen Variablen. *Beispiele für Schreibweisen*: $y = f(x)$, $x = g(t)$, $r = h(\varphi)$, ... oder auch $y = y(x)$, $x = x(t)$, $r = r(\varphi)$, ... In den zuletzt genannten Schreibweisen liegt eine – hier ausnahmsweise zulässige – Doppelbedeutung von Buchstaben vor. Der auf der linken Seite der Gleichung stehende Buchstabe bezeichnet die Elemente des Wertebereiches; dagegen bezeichnet dieser Buchstabe auf der rechten Seite die Zuordnungsvorschrift.

Geometrische Veranschaulichung von $y = y(x)$ erfolgt durch eine Kurve in einem kartesischen (x, y)-Koordinatensystem. Sie besteht aus den Punkten $(x, y(x))$.

13.3.1.2. Reelle Funktionen von mehreren reellen unabhängigen Variablen. *Beispiele für Schreibweisen*: $z = f(x, y)$, $U = g(x, y, z, t)$, $z = h(x_1, \ldots, x_n)$, ... oder auch $z = z(x, y)$, $U = U(x, y, z, t)$, $z = z(x_1, \ldots, x_n)$.

Definitionsbereich. Er ist bei einer reellen Funktion von n reellen unabhängigen Variablen einer Teilmenge aller n-Tupel $(x_1, \ldots, x_n)$, wobei $x_1, \ldots, x_n$ beliebige reelle Zahlen sind.

Geometrische Veranschaulichung. Im Falle $n = 2$ geschieht sie durch eine Fläche in einem kartesischen (x, y, z)-System. Diese besteht aus den Punkten $(x, y, z(x, y))$.

Niveaulinien. Dies sind im Falle $n = 2$ diejenigen Punkte (x, y) des Definitionsbereiches von $z = z(x, y)$, für die $z(x, y) = c$ (c = Konstante) gilt.

Niveauflächen. Im Falle $n = 3$ sind dies diejenigen Punkte

(x, y, z) des Definitionsbereiches von $U = U(x, y, z)$, für die $U(x, y, z) = c$ gilt. Es ist üblich, diese Sprechweise auch auf den Fall $n > 3$ zu übertragen.

13.3.1.3. Funktionen, deren Wertebereich aus Vektoren besteht, und Ortsvektoren, die von einer reellen unabhängigen Variablen abhängen. Sie liefern die *Parameterdarstellung* von Kurven. Durchläuft in $r = r(t) = x_k(t)e_k$ (Einsteinsche Summationsverabredung, vgl. 13.2.5.2.) die unabhängige Variable t – sie heißt im jetzigen Zusammenhang *Kurvenparameter* – den Definitionsbereich auf der Parameterachse, so durchlaufen die Endpunkte der zugehörigen Ortsvektoren die dargestellte orientierte Kurve.

Ortsvektoren, die von zwei reellen unabhängigen Variablen abhängen, liefern die Parameterdarstellung von Flächen. Durchläuft in $r = r(u, v) = x_k(u, v)e_k$ das Parameterpaar (u, v) in der Parameterebene den Definitionsbereich, so durchlaufen die Endpunkte der zugehörigen Ortsvektoren die dargestellte Fläche.

Ortsvektoren, die von drei reellen unabhängigen Variablen abhängen. Durch $r = r(u_1, u_2, u_3) = x_k(u_1, u_2, u_3)e_k$ werden krummlinige Koordinaten (u_1, u_2, u_3) eingeführt (vgl. 13.2.5.3.).

Vektorfelder. Durch $F = F(\mathrm{r}) = F(x_1, x_2, x_3) = F_k(x_1, x_2, x_3)e_k$ wird jedem Punkt (x_1, x_2, x_3) aus dem Definitionsbereich ein Vektor F zugeordnet, den man sich im Punkt (x_1, x_2, x_3) angeheftet denkt. Das Vektorfeld kann auch noch von der Zeit t abhängen: $F = F(r, t) = F(x_1, x_2, x_3, t)$.

13.3.1.4. Komplexe Funktionen einer unabhängigen Variablen. Komplexe Funktionen einer reellen unabhängigen Variablen liefern Parameterdarstellungen von Kurven in der Gaußschen Zahlenebene (vgl. 13.1.3.3.). Durchläuft in $z = z(t) = x(t) + \mathrm{i}y(t) = z(t)\,\mathrm{e}^{\mathrm{i}\varphi(t)}$ der Kurvenparameter t den Definitionsbereich, so durchläuft die zugehörige komplexe Zahl z in der Gaußschen Zahlenebene die dargestellte orientierte Kurve.

Komplexe Funktionen einer komplexen unabhängigen Variablen. Durch $w = w(z)$ $(w = u + \mathrm{i}v,\ z = x + \mathrm{i}y)$ und damit durch $u = u(x,y)$, $v = v(x,y)$ wird eine Abbildung vermittelt von dem in der Gaußschen z-Ebene liegenden Definitionsbereich auf den in der Gaußschen w-Ebene liegenden Wertebereich.

13.3.1.5. Funktionale, die in der Variationsrechnung behandelt werden. Ist in 13.3.1. die Menge M_2 gleich dem Körper der reellen oder komplexen Zahlen, so nennt man den vorliegenden Operator auch *Funktional*. Ein solches wird beispielsweise durch das *Wirkungsintegral* $S = \int_a^b L(t, q_1, \ldots, q_f, \dot{q}_1, \ldots, \dot{q}_f)\,\mathrm{d}t$ gegeben. (L = *Lagrange-Funktion*, t = Zeit, $q_1, \ldots, q_f$ = *verallgemeinerte Koordinaten*, f = Freiheitsgrade, $\dot{q}_1, \ldots, \dot{q}_f$ = *verallgemeinerte Geschwindigkeiten*.) Der Definitionsbereich besteht aus f-tupeln $(q_1(t), \ldots, q_f(t))$ *von reellen Funktionen* $q_\nu(t)$ $(\nu = 1, \ldots, f)$ der einen reellen Variablen t. Sie heißen *Vergleichsfunktionen*. Es ist auch möglich, daß in der obigen Formel noch höhere Ableitungen nach t auftreten oder daß die Vergleichsfunktionen von mehreren unabhängigen Variablen abhängen. Im letzteren Fall ist das Integral über das t-Integrationsintervall durch das entsprechende Integral über einen mehrdimensionalen Bereich zu ersetzen.

13.3.1.6. Integraloperatoren. Bsp. 13.3.1–1. Das Anfangswertproblem für die gesuchte Funktion $y = y(x)$ bestehend aus der gewöhnlichen Differentialgleichung erster Ordnung $y' = f(x, y)$ und der Anfangsbedingung $y(x_0) = y_0$ ist äquivalent zur (i. allg. nichtlinearen) *Integralgleichung* (diese Sprechweise, da die unbekannte Funktion im Integranden steht) $y = T[y]$, wobei der *Integraloperator* T durch $T[y] = y_0 + \int_{x_0}^{x} f[t, y(t)]\,\mathrm{d}t$ gegeben ist.

Bsp. 13.3.1–2. Durch $T[\varphi] = \int_a^b K(x,y)\varphi(y)\mathrm{d}y$ ist ein *linearer Integraloperator* gegeben. $K(x, y)$ heißt Kern. *Integralgleichungen erster* bzw. *zweiter Art* haben die Struktur $T[\varphi] + f(x) = 0$ bzw. $T[\varphi] + f(x) = \varphi(x)$. Sie heißen im Fall $f(x) = 0$ *homogen*, sonst *inhomogen* (vgl. Bsp. 13.1.5–6, 13.1.5–7).

13.3.1.7. Zufallsgröße X ist eine Funktion $X = X(\omega)$, deren Definitionsbereich gleich der Menge Ω der elementaren Ereignisse eines Wahrscheinlichkeitsraumes $[\Omega, B, P]$ (vgl. 13.1.6.6.) ist und deren Wertebereich aus reellen Zahlen besteht. Weiterhin gehört zum Begriff der Zufallsgröße eine ihr zugeordnete *Verteilungsfunktion* $F = F(x)$, deren Definitionsbereich die Menge der reellen Zahlen ist. Zur Bestimmung des Funktionswertes $F(x)$ betrachtet man alle diejenigen $\omega \in \Omega$, für die $-\infty < X(\omega) < x$ gilt, setzt voraus, daß für jedes x die von diesen ω gebildete Teilmenge von Ω zum System B von $[\Omega, B, P]$ gehört, und versteht unter $F(x)$ die Wahrscheinlichkeit desjenigen Ereignisses, das durch die genannte Teilmenge von Ω dargestellt wird. Man schreibt $F(x) = P(X < x)$. Die Zufallsgröße X heißt *diskret*, wenn sie höchstens abzählbar viele Werte annehmen kann; sie heißt *stetig*, wenn es eine Funktion $f = f(x)$ – *Wahrscheinlichkeitsdichte* genannt – gibt mit $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$.

Bsp. 13.3.1–3. Interessiert man sich bei n-maligen voneinander unabhängigen Durchführungen eines Versuches dafür, ob ein Ereignis A mit der gegebenen Wahrscheinlichkeit p eintritt, so ist die Anzahl der Durchführungen, bei denen A eintritt, eine diskrete Zufallsgröße X mit dem Wertebereich $0, 1, \ldots, n$. X ist *binomialverteilt* mit den Parametern p und n, d.h. es liegen die *Einzelwahrscheinlichkeiten* $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ $(k = 0,\ 1,\ \ldots,\ n)$ vor, und damit ist $F(x) = \sum \binom{n}{k} p^k (1 - p)^{n-k}$, wobei über diejenigen $k = 0,\ 1,\ \ldots,\ n$ zu summieren ist, die kleiner als x sind.

Bsp. 13.3.1–4. X heißt *normalverteilt* mit den Parametern μ und σ^2, (Schreibweise: $N(\mu, \sigma^2)$-verteilt), wenn X die Wahrscheinlichkeitsdichte hat ($\exp x = \mathrm{e}^x$)

$$f(x) = (\sqrt{2\pi}\,\sigma)^{-1} \exp[(x - \mu)^2/(2\sigma^2)].$$

13.3.2. Banachscher Fixpunktsatz

Die Mengen M_1 und M_2 aus 13.3.1. seien nunmehr einander gleich und metrisch (Bezeichnung $M_1 = M_2 = M$) und $D(F) \subseteqq M$ sei ein *vollständiger* metrischer Raum, d.h., jede Cauchyfolge aus $D(F)$ sei eine konvergente Folge, deren Grenzwert zu $D(F)$ gehört (vgl. 13.1.4.1., ϱ ist Metrik in M). Die Funktion F bilde $D(F)$ in sich ab, d.h., der Wertevorrat von F sei eine Teilmenge des Definitionsbereiches $D(F) \subseteqq M$. Weiterhin sei F ein *Kontraktionsoperator*, d.h., es soll eine Zahl q mit $0 < q < 1$ derart geben, daß für alle u, v aus $D(F)$ die Ungleichung $\varrho(Fu, Fv) \leqq q\varrho(u, v)$ gilt. Dann besitzt F genau einen *Fixpunkt* u_{Fix}, d.h., es gibt genau ein Element u_{Fix} in $D(F)$, für das die *Fixpunktgleichung* $F(u_{\text{Fix}}) = u_{\text{Fix}}$ gilt. Entnimmt man $D(F)$ irgendein Element

u_0 und bildet man gemäß der *Iterationsvorschrift* $u_{n+1} = F(u_n)$ $(n = 0, 1, 2, \ldots)$ die Folge $u_0, u_1, u_2, \ldots$, so ist $\lim_{n\to\infty} u_n = u_{\text{Fix}}$. Schließlich gilt für die Entfernung der n-ten Näherung u_n vom Fixpunkt u_{Fix} die Ungleichheit $\varrho(u_n, u_{\text{Fix}}) \leqq (q^n/(1-q))\varrho(u_1, u_0)$; man spricht hier von einer *Fehlerabschätzung*. Wenn die „Kugel" K, bestehend aus allen u mit $\varrho(u, u_1) \leqq (q/(1-q))\varrho(u_1, u_0)$ zu M gehört, so kann man durch Wahl von K als $D(F)$ sichern, daß der Wertevorrat von F zu $D(F)$ gehört. In diesem Fall wird die Aussage $u_0 \in D(F)$ durch $u_0 \in M$ ersetzt.

13.3.2.1. Eine nichtlineare Gleichung mit einer Unbekannten. Die Funktion $y = f(x)$ habe für $a \leqq x \leqq b$ eine stetige Ableitung $f'(x)$ mit $|f'(x)| \leqq q < 1$. Ist weiterhin bekannt, daß der Wertevorrat im Definitionsbereich enthalten ist (vgl. 13.3.2.), so hat die Gleichung $x = f(x)$ in $a \leqq x \leqq b$ genau eine Lösung x_{Fix}, die Grenzwert einer Zahlenfolge $x_0, x_1, x_2, \ldots$ ist, die ausgehend von der nullten Näherung x_0 aus dem Intervall $a \leqq x \leqq b$ gemäß der Iterationsvorschrift $x_{n+1} = f(x_n)$ $(n = 0, 1, \ldots)$ gebildet wird. Es gilt die Fehlerabschätzung $|x_n - x_{\text{Fix}}| \leqq \frac{q^n}{1-q}|x_1 - x_0|$.

Bsp. 13.3.2–1. Die Gleichung $x = \tan x$ hat unendlich viele Lösungen; man erkennt dies unmittelbar an den Kurvenbildern der Funktionen $y = x$ und $y = \tan x$. Gesucht sei die kleinste positive Lösung. An der Figur erkennt man, daß sie im Intervall $\pi < x < (3/2)\pi$ liegt. Die Theorie ist nicht unmittelbar anwendbar, da dort $(\tan x)' > 1$ ist. Löst man die Gleichung $y = \tan x$ im Intervall $\pi/2 < x < (3/2)\pi$ nach x auf, so ergibt sich $x = \pi + \arctan y$. Die gesuchte Lösung x_{Fix} genügt also der Gleichung $x_{\text{Fix}} = \pi + \arctan x_{\text{Fix}}$. Jetzt sind im interessierenden Intervall $\pi \leqq x \leqq (3/2)\pi$ die Voraussetzungen des Banachschen Fixpunktsatzes erfüllt.

Bsp. 13.3.2.–2. Newtonsches Verfahren. Ist x_n die n-te Näherung der Lösung x von $f(x)=0$, so ersetzt man zur Berechnung von x_{n+1} die Kurve $y = f(x)$ durch ihre Tangente an der Stelle x_n und berechnet den Schnittpunkt mit der x-Achse: $x_{n+1} = x_n - f(x_n)/f'(x_n)$. Diese Iterationsvorschrift gehört zu der zu $f(x) = 0$ äquivalenten Fixpunktgleichung $x = g(x)$ mit $g(x) = x - f(x)/f'(x)$.

13.3.2.2. Der Satz von Picard-Lindelöf. Für die Integralgleichung (Fixpunktgleichung)

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,dt$$

(vgl. Bsp. 13.3.1.–1.) sind die Voraussetzungen des Banachschen Fixpunktsatzes erfüllt, falls man als Definitionsbereich die Menge der stetigen Funktionen $y = y(x)$ $(|x - x_0| \leqq c)$ mit $\|y\| = \max|y(x)|$ wählt, die Stetigkeit von $f = f(x,y)$ für $|x - x_0| \leqq a$, $|y - y_0| \leqq b$ voraussetzt und darüber hinaus dort $|f(x,y) - f(x,\bar{y})| < L|y - \bar{y}|$ (*Lipschitzbedingung*), $|f(x,y)| \leqq K$ fordert mit $0 < c < a$, $cK < b$, $cL < 1$.

13.3.3. Definition der Differenzierbarkeit

Grenzwert. Sind die Mengen M_1 und M_2 aus 13.3.1. metrische Räume (vgl. 13.1.4.1.), so schreibt man $\lim_{u\to u_0} F(u) = A$ (u_0 ist Häufungspunkt von $D(F) \subseteq M_1$, $A \in M_2$), wenn zu jeder (noch so kleinen) reellen Zahl $\varepsilon > 0$ eine reelle Zahl $\delta > 0$ derart existiert, daß für alle u aus $D(F)$ mit $0 < \varrho_1(u, u_0) < \delta$ (ϱ_1 ist Metrik in M_1) stets $\varrho_2(F(u), A) < \varepsilon$ (ϱ_2 ist Metrik in M_2) folgt.

Stetigkeit. Die Funktion F heißt im Punkt u_0 von $D(F)$ stetig, wenn $\lim_{u\to u_0} F(u) = F(u_0)$ gilt. Sie heißt stetig, wenn sie in jedem Punkt ihres Definitionsbereiches stetig ist.

Banach-Raum stetiger linearer Operatoren. Im Vektorraum M linearer Operatoren aus 13.1.5.1. sei jetzt der gemeinsame Definitionsbereich D und der dortige Vektorraum G_2 ein Banach-Raum (vgl. 13.1.4.2.). Die Teilmenge aller *stetigen* linearen Operatoren aus M ist gleich der Teilmenge aller *beschränkten* linearen Operatoren; hierbei heißt ein linearer Operator T beschränkt, falls es eine positive Zahl C derart gibt, daß für alle u aus D stets $\|Tu\| \leqq C\|u\|$ gilt. Die kleinste derartige Zahl C heißt Norm von T (Schreibweise: $\|T\|$). Dieser Name ist berechtigt, denn mit der Norm $\|T\|$ bildet die Menge dieser stetigen linearen Operatoren einen *Banach-Raum*. Ist in der assoziativen Algebra linearer Operatoren aus 13.1.5.2. der gemeinsame Definitionsbereich ein Banach-Raum, so gilt in der Teilmenge der *stetigen* linearen Operatoren über die Gesetze eines Banach-Raumes hinaus noch $\|T_1T_2\| \leqq \|T_1\|\,\|T_2\|$ und die Norm des die identische Abbildung vermittelnden Operators ist gleich 1. Dies wird in der Sprechweise *Banach-Algebra stetiger linearer Operatoren* zusammengefaßt.

(Fréchet-)Differenzierbarkeit. Sind die Mengen M_1 und M_2 aus 13.3.1. Banach-Räume, so heißt die Funktion F an der Stelle $u_0 \in D(F)$ *Fréchet-differenzierbar*, wenn es einen stetigen *linearen* Operator T – er heißt Ableitung $F'(u_0)$ – derart gibt, daß im Sinne einer lokalen linearen Approximation die Funktionswertdifferenz $F(u_0 + h) - F(u_0)$ bis auf eine *Größe höherer Ordnung* in der Umgebung von u_0 durch Th angegeben werden kann. Es muß also $F(u_0 + h) - F(u_0) = Th + o(\|h\|)$ sein, wobei für $o(\|h\|)$ (lies: klein o von $\|h\|$) $\lim_{\|h\|\to 0} (\|h\|^{-1}o(\|h\|) = 0$ gilt. (Der Grenzwert null ist hierbei das Nullelement des Banach-Raumes M_2.) Th wird *Fréchet-)-Differential* genannt. Die Funktion F heißt (schlechthin) differenzierbar, wenn sie an jeder Stelle ihres Definitionsbereiches differenzierbar ist.

13.3.3.1. Differenzierbarkeit von reellen Funktionen einer reellen unabhängigen Variablen. Die Ableitung von $y = f(x) = y(x)$ an der Stelle $x = x_0$ – Schreibweisen: $f'(x_0) = y'(x_0) = (f'(x))_{x=x_0} = (y'(x))_{x=x_0} = \frac{df}{dx}(x_0) = \frac{dy}{dx}(x_0)$ – ist eine reelle Zahl; sie ist gleich dem Anstieg der Tangente an die Kurve $y = y(x)$ an der Stelle $x = x_0$. Für das *Differential* $dy = df$ (*der abhängigen Variablen y*) gilt $dy = y'(x_0)dx$, wobei $dx = h = x - x_0$ das *Differential der unabhängigen Variablen x* ist. Gemäß 13.3.3. erhält man für die Funktionswertdifferenz $\Delta y = y(x_0 + dx) - y(x_0) = dy + o(|dx|)$. Ist $y = y(x)$ an jeder Stelle des Definitionsbereiches differenzierbar, so schreibt man für die Ableitung $y' = y'(x) = dy/dx = (d/dx)y(x) = y_{,x}$. Ableitungen nach der Zeit t werden oft durch einen Punkt gekennzeichnet: $\dot{x}(t) = dx/dt$.

13.3.3.2. Partielle Ableitung. Ist die Funktion $U = U(x, y, z)$ bei festem $y = y_0$ und $z = z_0$ in bezug auf x an der Stelle $x = x_0$ differenzierbar, so bezeichnet man die Ableitung mit $(\partial U/\partial x)$ $(x_0, y_0, z_0) = U_{,x}(x_0, y_0, z_0)$ und spricht von der *partiellen Ableitung* von U nach x an der Stelle (x_0, y_0, z_0). Analog definiert man $\partial U/\partial y = U_{,y}$ und $\partial U/\partial z = U_{,z}$ an der Stelle (x_0, y_0, z_0). Sind Mißverständnisse ausgeschlossen, so können in $U_{,x}$, $U_{,y}$, $U_{,z}$ die Kommata weggelassen werden.

13.3.3.3. (Totale) Differenzierbarkeit von reellen Funktionen mehrerer reeller unabhängiger Variabler. Ist U

$= U(x,y,z)$ an der Stelle (x_0, y_0, z_0) Fréchet-differenzierbar (vgl. 13.3.3.) – Sprechweise: *total differenzierbar* – so gilt für die Funktionswertdifferenz $\Delta U = U(x_0 + \mathrm{d}x, y_0 + \mathrm{d}y, z_0 + \mathrm{d}z) - U(x_0, y_0, z_0) = \mathrm{d}U + o(|\mathrm{d}\boldsymbol{r}|)$ mit dem *vollständigen Differential* $\mathrm{d}U = \mathbf{grad}\, U \cdot \mathrm{d}\boldsymbol{r}$; hierbei ist $\mathrm{d}\boldsymbol{r} = \mathrm{d}x\,\boldsymbol{e}_x + \mathrm{d}y\,\boldsymbol{e}_y + \mathrm{d}z\,\boldsymbol{e}_z$ das *vektorielle Differential* des Ortsvektors $\boldsymbol{r} = x\,\boldsymbol{e}_x + y\,\boldsymbol{e}_y + z\,\boldsymbol{e}_z$ an der Stelle (x_0, y_0, z_0) (also: $\mathrm{d}x = x - x_0$, $\mathrm{d}y = y - y_0$, $\mathrm{d}z = z - z_0$) und die Fréchet-Ableitung von U ist ein Vektor, nämlich der auf den Niveauflächen von U (vgl. 13.3.1.2.) jeweils senkrecht stehende Gradient $\mathbf{grad}\, U = (\partial U/\partial x)\boldsymbol{e}_x + (\partial U/\partial y)\boldsymbol{e}_y + (\partial U/\partial z)\boldsymbol{e}_z$ (vgl. 13.2.6.7.). Damit ist $\mathrm{d}U = (\partial U/\partial x)\mathrm{d}x + (\partial U/\partial y)\mathrm{d}y + (\partial U/\partial z)\mathrm{d}z$. Analog ist für die Funktion $z = z(x_1, \ldots, x_n)$ die Funktionswertdifferenz $\Delta z = \mathrm{d}z + o(|\mathrm{d}\boldsymbol{r}|)$ mit $\mathrm{d}z = \mathbf{grad}\, z \cdot \mathrm{d}\boldsymbol{r}$, $\mathbf{grad}\, z = z_{,x_k}\boldsymbol{e}_{x_k}$ (Einsteinsche Summationsverabredung mit $k = 1, 2, \ldots, n$), $\mathrm{d}\boldsymbol{r} = \mathrm{d}x_k\boldsymbol{e}_{x_k}$; und damit lautet das *vollständige Differential* $\mathrm{d}z = z_{,x_1}\mathrm{d}x_1 + \ldots + z_{,x_n}\mathrm{d}x_n$.

13.3.3.4. Tangentenvektoren an Raumkurven. Die Fréchet-Ableitung der Parameterdarstellung $\boldsymbol{r}(t) = x_k(t)\boldsymbol{e}_k$ (vgl. 13.3.1.3.) einer Raumkurve ist ein Vektor, nämlich (häufige Bezeichnung mit einem Punkt, da t oft die Rolle der Zeit spielt) $\dot{\boldsymbol{r}}(t) = \dot{x}_k(t)\boldsymbol{e}_k$, er wird im Kurvenpunkt mit dem Ortsvektor $\boldsymbol{r}(t)$ angeheftet und weist in Richtung der Tangente an die Kurve im genannten Kurvenpunkt. Für die Ortsvektordifferenz $\Delta\boldsymbol{r} = \boldsymbol{r}(t_0 + \mathrm{d}t) - \boldsymbol{r}(t_0)$ gilt $\Delta\boldsymbol{r} = \mathrm{d}\boldsymbol{r} + o(|\mathrm{d}t|)$ mit dem *vektoriellen Kurvendifferential* $\mathrm{d}\boldsymbol{r} = \dot{\boldsymbol{r}}(t_0)\mathrm{d}t(\mathrm{d}t = t - t_0)$.

13.3.3.5. Tangentenvektoren an Flächen. Die Fréchet-Ableitung der Parameterdarstellung $\boldsymbol{r}(u, v) = x_k(u, v)\boldsymbol{e}_k$ (vgl. 13.3.1.3.) einer Fläche ist durch ein Vektorpaar gegeben, nämlich $\partial\boldsymbol{r}/\partial u = (\partial x_k/\partial u)\boldsymbol{e}_k$ und $\partial\boldsymbol{r}/\partial v = (\partial x_k/\partial v)\boldsymbol{e}_k$. Für die Ortsvektordifferenz $\Delta\boldsymbol{r} = \boldsymbol{r}(u_0 + \mathrm{d}u, v_0 + \mathrm{d}v) - \boldsymbol{r}(u_0, v_0)$ gilt $\Delta\boldsymbol{r} = \mathrm{d}\boldsymbol{r} + o\left(\sqrt{(\mathrm{d}u)^2 + (\mathrm{d}v)^2}\right)$ mit dem *vektoriellen Kurvendifferential* auf der Fläche $\mathrm{d}\boldsymbol{r} = (\partial\boldsymbol{r}/\partial u)\mathrm{d}u + (\partial\boldsymbol{r}/\partial v)\mathrm{d}v$ $(\mathrm{d}u = u - u_0, \mathrm{d}v = v - v_0)$, das man sich im Flächenpunkt mit dem Ortsvektor $\boldsymbol{r}(u_0, v_0)$ angeheftet denkt und das dann in der dortigen Tangentialebene der Fläche liegt.

13.3.3.6. Vektorielles Differential in krummlinigen Koordinaten. Die Fréchet-Ableitung von $\boldsymbol{r}(u_1, u_2, u_3) = x_k(u_1, u_2, u_3)\,\boldsymbol{e}_k$ (vgl. 13.3.1.3.) ist durch ein Vektortripel gegeben, nämlich $\partial\boldsymbol{r}/\partial u_l = (\partial x_k/\partial u_l)\boldsymbol{e}_k (l = 1, 2, 3)$. Für die Ortsvektordifferenz $\Delta\boldsymbol{r} = \boldsymbol{r}(u_{10} + \mathrm{d}u_1, u_{20} + \mathrm{d}u_2, u_{30} + \mathrm{d}u_3) - \boldsymbol{r}(u_{10}, u_{20}, u_{30})$ gilt $\Delta\boldsymbol{r} = \mathrm{d}\boldsymbol{r} + o\left(\sqrt{(\mathrm{d}u_1)^2 + (\mathrm{d}u_2)^2 + (\mathrm{d}u_3)^2}\right)$ mit dem *vektoriellen Differential* $\mathrm{d}\boldsymbol{r} = (\partial\boldsymbol{r}/\partial u_l)\mathrm{d}u_l(\mathrm{d}u_l = u_l - u_{l0}, l = 1, 2, 3)$.

13.3.3.7. (Totale) Differenzierbarkeit in Vektorfeldern. Ist das Vektorfeld $\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r}) = F_k(x_1, x_2, x_3)\boldsymbol{e}_k$ (vgl. 13.3.1.3.) an der Stelle (x_{10}, x_{20}, x_{30}) Fréchet-differenzierbar, so gilt für die vektorielle Funktionswertdifferenz $\Delta\boldsymbol{F} = \boldsymbol{F}(\boldsymbol{r} + \mathrm{d}\boldsymbol{r}) - \boldsymbol{F}(\boldsymbol{r}) = \mathrm{d}\boldsymbol{F} + o(|\mathrm{d}\boldsymbol{r}|)$ mit dem vektoriellen vollständigen Differential $\mathrm{d}\boldsymbol{F} = (\partial\boldsymbol{F}/\partial x_l)\mathrm{d}x_l = (\partial F_k/\partial x_l)\mathrm{d}x_l\boldsymbol{e}_k = (\partial_l F_k \mathrm{d}x_l)\boldsymbol{e}_k = F_{k,l}\mathrm{d}x_l\boldsymbol{e}_k$. Die Fréchet-Ableitung ist also ein *Tensor zweiter Stufe*, der relativ zum vorliegenden Koordinatensystem durch die quadratische Matrix $(F_{k,l})$ dargestellt wird. Sie bildet die Koordinaten $\mathrm{d}x_l$ von $\mathrm{d}\boldsymbol{r}$ in die Koordinaten $\mathrm{d}F_k$ von $\mathrm{d}\boldsymbol{F}$ ab (vgl. 13.2.6.2.).

13.3.3.8. Differentiation komplexer Funktionen einer reellen unabhängigen Variablen. Ist $z(t) = x(t) + \mathrm{i}y(t)$ die Parameterdarstellung einer Kurve in der Gaußschen Zahlenebene (vgl. 13.3.1.4.), und interpretiert man die komplexe Zahl $\dot{z}(t) = \dot{x}(t) + \mathrm{i}\dot{y}(t)$ als Vektor in der Gaußschen Zahlenebene, der am Kurvenpunkt $z(t)$ angeheftet ist, so weist dieser *Vektor in Richtung der Tangente* der Kurve im genannten Kurvenpunkt.

13.3.3.9. Differentiation komplexer Funktionen einer komplexen unabhängigen Variablen. Ist die Funktion $w(z) = u(x,y) + \mathrm{i}v(x,y)$ $(z = x + \mathrm{i}y, w = u + \mathrm{i}v)$ in einer Umgebung von $z_0 = x_0 + \mathrm{i}y_0$ (d. h. für alle z mit $|z - z_0| < \varepsilon, \varepsilon > 0$ beliebig [klein]) Fréchet-differenzierbar – Sprechweise: $w(z)$ ist *holomorph* (auch *regulär* oder *regulär-analytisch*) an der Stelle $z = z_0$ – so gilt für die Funktionswertdifferenz $\Delta w = w(z_0 + \mathrm{d}z) - w(z_0) = \mathrm{d}w + o(|\mathrm{d}z|)$ mit dem Differential $\mathrm{d}w = w'(z_0)\mathrm{d}z(\mathrm{d}z = z - z_0)$, wobei $w'(z) = (\partial u/\partial x) + \mathrm{i}(\partial v/\partial x) = (1/\mathrm{i})((\partial u/\partial y) + \mathrm{i}(\partial v/\partial y))$ ist und daher die *Cauchy-Riemannschen Differentialgleichungen* $\partial u/\partial x = \partial v/\partial y$, $\partial u/\partial y = -\partial v/\partial x$ gelten. Die durch holomorphe Funktionen $w(z)$ vermittelten Abbildungen sind im Falle $w'(z) \neq 0$ *konform* (*winkeltreu*), d. h. Kurven der z-Ebene, die sich unter dem orientierten Winkel α schneiden, gehen in Kurven der w-Ebene über, die sich ebenfalls unter dem orientierten Winkel α schneiden (Beispiele siehe 13.4.7.4.).

13.3.3.10. Die Variation des Wirkungsintegrals. Unter der *Variation* δS des Wirkungsintegrals S aus 13.3.1.5. versteht man das Fréchet-Differential von S. Es ist

$$\delta S = \sum_{k=1}^{f}\left\{\int_a^b\left(\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\left[\frac{\partial L}{\partial \dot{q}_k}\right]\right)\delta q_k \mathrm{d}t + \left[\frac{\partial L}{\partial \dot{q}_k}\delta q_k\right]_{t=a}^{t=b}\right\}$$

mit der *Variation der Vergleichsfunktionen* $\delta q_k = \delta q_k(t) = \tilde{q}_k(t) - q_k(t)$. In der Variationsrechnung interessiert man sich für diejenigen Funktionen f-Tupel $(q_1(t), \ldots, q_f(t))$, für die das Wirkungsintegral einen relativen Extremwert annimmt. Eine notwendige Bedingung ist $\delta S = 0$. Hieraus folgen einerseits das System der *Eulerschen Differentialgleichungen der Variationsrechnung* (*Lagrangesche Gleichungen zweiter Art*) $(\mathrm{d}/\mathrm{d}t)(\partial L/\partial\dot{q}_k) - \partial L/\partial q_k = 0$ $(k = 1, \ldots, f)$ und andererseits – falls von der Aufgabenstellung her keine Randbedingungen gefordert werden – die *natürlichen Randbedingungen* $[\partial L/\partial\dot{q}_k]_{t=a} = 0$ und $[\partial L/\partial\dot{q}_k]_{t=b} = 0$ $(k = 1, \ldots, f)$.

13.3.4. Taylor-Entwicklung

13.3.4.1. Ableitung zweiter Ordnung. Ist die Funktion F aus 13.3.3. mit den Funktionswerten $F(u) \in M_2$ (M_2 = Banach-Raum) in ihrem Definitionsbereich $D(F) \subseteqq M_1$ (M_1 = Banach-Raum) Fréchet-differenzierbar, existiert also für alle $u \in D(F)$ die Ableitung $F'(u)$, so ist dadurch eine Funktion F' gegeben. Ist diese ihrerseits differenzierbar, so sagt man F sei zweimal differenzierbar und schreibt für die Ableitung von F' an der Stelle $u = u_0$ das Symbol $F''(u_0)$. Das bedeutet bei Benutzung der Schreibweise $B = F''(u_0)$ wegen 13.3.3. $F'(u_0 + h) - F'(u_0) = Bh + o(\|h\|)$, $(h \in M_1)$. Hierbei ist B ein stetiger linearer Operator, der die Elemente $h \in M_1$ abbildet in den Wertebereich von F', dessen Elemente definitionsgemäß ihrerseits stetige lineare Operatoren T sind, wobei Tk mit $k \in M_1$ zum Raum M_2 gehört. Also kann wegen $Tk = (Bh)k$ das B als Operator aufgefaßt werden, der das Elementepaar (h, k) $(h \in M_1, k \in M_1)$ jeweils in ein Element von M_2 abbildet. B ist dann ein stetiger *bilinearer Operator* $B = B(h, k)$, d.h. B ist sowohl bezüglich h als auch bezüglich k ein stetiger linearer Operator, und es gibt eine Zahl $C > 0$ mit $\|B(h,k)\| \leqq C\|h\|\,\|k\|$.

Bsp. 13.3.4–1. Die Fréchet-Ableitung der Funktion $U = U(x_1, x_2, x_3)$ wird durch das Gradientenfeld $\mathbf{grad}\ U = \frac{(\partial U)}{(\partial x_\mu)}\boldsymbol{e}_\mu$ (vgl. 13.3.3.3.) gegeben. Folglich ist wegen

13.3.3.7. die zweite Fréchet-Ableitung von U ein Tensorfeld zweiter Stufe, das relativ zum vorliegenden kartesischen Koordinatensystem (x_1, x_2, x_3) durch die quadratische Matrix $\left(\frac{\partial^2 U}{\partial x_\mu \partial x_\nu}\right)$ dargestellt wird. Dieser Tensor kann als bilinearer Operator aufgefaßt werden, der dem Vektorpaar (h, k) $(h = h_\mu e_\mu,\ k = k_\nu e_\nu)$ die Zahl $\frac{\partial^2 U}{\partial x_\mu \partial x_\nu} h_\mu k_\nu$ zuordnet.

13.3.4.2. Ableitung höherer Ordnung. Im Sinne von 13.3.4.1. fortfahrend werden die weiteren Ableitungen F''', F'''', ..., $F^{(n)}$ eingeführt. $F^{(n)}(u_0)$ ist dann ein *stetiger n-linearer Operator N*, d. h. ein Operator N mit den Funktionswerten $N(w_1, \ldots, w_n)$, $(w_1 \in M_1, \ldots, w_n \in M_1)$, der bezüglich jedes w_ν $(\nu = 1, \ldots, n)$ jeweils ein stetiger linearer Operator ist und für den es ein $C > 0$ mit $\|N(w_1, \ldots, w_n)\| < C\|w_1\| \, \|w_2\| \ldots \|w_n\|$ gibt.

Bsp. 13.3.4.-2. Im Bsp. 13.3.4-1 ist die dritte Fréchet-Ableitung von U ein Tensor dritter Stufe, der relativ zum vorliegenden Koordinatensystem durch $\frac{\partial^3 U}{\partial x_\mu \partial x_\nu \partial x_\varkappa}$ $(\mu = 1, 2, 3,$ $\nu = 1, 2, 3,$ $\varkappa = 1, 2, 3)$ angegeben werden kann. Es liegt damit ein 3-linearer Operator N vor mit $N(h, k, l) = \frac{\partial^3 U}{\partial x_\mu \partial x_\nu \partial x_\varkappa} h_\mu k_\nu l_\varkappa$ $\quad (h = h_\mu e_\mu,\ k = k_\nu e_\nu,\ l = l_\varkappa e_\varkappa)$.

13.3.4.3. Angabe der Taylor-Formel. Mit den Bezeichnungen aus 13.3.4.1. und 13.3.4.2. lautet die Taylor-Formel:

$$F(u_0 + h) = F(u_0) + F'(u_0)h + (1/2!)F''(u_0)(h, h) + \ldots + (1/n!)F^{(n)}(u_0)(\underbrace{h, \ldots, h}_{n \text{mal}}) + R_{n+1}$$

(es ist $\nu!$ – gelesen ν-Fakultät – durch $0! = 1$, $\nu! = 1 \cdot 2 \cdot \ldots \cdot \nu$ definiert) mit $\|R_{n+1}\| \leqq M \frac{\|h\|^{n+1}}{(n+1)!}$, wobei die Zahl M größer und höchstens gleich allen Normen sein muß, die mit $0 \leqq t \leqq 1$ durch $\|F^{(n+1)}(u_0 + th)\|$ angegeben werden. In dem Fall $n = 0$ ergibt sich der *Mittelwertsatz der Differentialrechnung* $\|F(u_0 + h) - F(u_0)\| \leqq M\|h\|$ mit $M \geqq \|F'(u_0 + th)\|$, $0 \leqq t \leqq 1$.

13.3.4.4. Taylor-Entwicklung holomorpher Funktionen. Für jede an der Stelle $z = z_0$ holomorphe Funktion $w = w(z)$ (vgl. 13.3.3.9.) existieren dort alle Ableitungen höherer Ordnung, und in der Taylor-Formel (vgl. 13.3.4.3.) $w(z) = w(z_0 + h) = w(z_0 + (z - z_0)) = \sum_{\nu=0}^{n} \frac{1}{\nu!} w^{(\nu)}(z_0)(z - z_0)^\nu + R_{n+1}$ gilt für alle z mit $|z - z_0| < r$ (r = Konvergenzradius) $\lim_{n\to\infty} R_{n+1} = 0$, d.h. es gilt

$$w(z) = \lim_{n\to\infty} \sum_{\nu=0}^{n} \frac{1}{\nu!} w^{(\nu)}(z_0)\,(z - z_0)^\nu \text{ für } |z - z_0| < r.$$

Man sagt hierfür, daß $w(z)$ an der Stelle $z = z_0$ in eine *Taylor-Reihe* (*Potenzreihe*) entwickelbar ist und schreibt

$$w(z) = \sum_{\nu=0}^{\infty} \frac{1}{\nu!} w^{(\nu)}(z_0)(z - z_0)^\nu$$

oder auch $w(z) = w(z_0) + w'(z_0)(z - z_0) + (1/2)w''(z_0)(z - z_0)^2 + (1/3!)w'''(z_0)(z - z_0)^3 + \ldots$.

Die Entwicklung ist gültig (und die Taylor-Reihe *konvergiert*) für alle z mit $|z - z_0| < r$; für alle z mit $|z - z_0| > r$ *divergiert* die Taylor-Reihe und die Entwicklung ist nicht gültig. Aussagen darüber, ob in einigen oder allen Punkten des *Randes des Konvergenzkreises* $|z - z_0| = r$ die Taylor-Entwicklung gültig ist, können allgemein nicht gemacht werden. Zur Berechnung von r kann man einen Kreis mit dem Mittelpunkt z_0 vom Radius null an solange „aufblähen" bis er an eine – z_0 am nächsten gelegene – Stelle anstößt, wo die Holomorphie von $w(z)$ verletzt ist, mit anderen Worten: bis zur nächstgelegenen Singularität von $w = w(z)$.

13.3.4.5. Rechnen mit Potenzreihen. Addition.

$$\sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu + \sum_{\nu=0}^{\infty} b_\nu(z - z_0)^\nu = \sum_{\nu=0}^{\infty} (a_\nu + b_\nu)(z - z_0)^\nu.$$

Der Konvergenzradius der Ergebnisreihe ist mindestens so groß wie der kleinere Radius der beiden Summanden.

Multiplikation. $\left(\sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu\right)\left(\sum_{\nu=0}^{\infty} b_\nu(z - z_0)^\nu\right) = \sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu$ mit $c_\nu = \sum_{k=0}^{\nu} a_k b_{\nu-k}$. Der Konvergenzradius der Ergebnisreihe ist mindestens so groß wie der kleinere Radius der beiden Faktoren.

Division. *Erster Fall*: Die Nennerpotenzreihe ist für $z = z_0$ ungleich null.

$$\left(\sum_{\nu=0}^{\infty} b_\nu(z - z_0)^\nu\right)^{-1} \sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu = \sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu,$$

wobei wegen $\sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu = \left(\sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu\right) \left(\sum_{\nu=0}^{\infty} b_\nu(z - z_0)^\nu\right) = \sum_{\nu=0}^{\infty}\left(\sum_{k=0}^{\nu} c_k b_{\nu-k}\right)(z - z_0)^\nu$ sich durch *Koeffizientenvergleich* das Gleichungssystem $a_\nu = \sum_{k=0}^{\nu} c_k b_{\nu-k}$ $(\nu = 0, 1, \ldots)$ ergibt, aus dem die unbekannten Koeffizienten $c_0, c_1, \ldots$ der Reihe nach bestimmt werden können: $c_0 = a_0/b_0$, $c_1 = (b_0 a_1 - a_0 b_1)/b_0^2$, ...

Zweiter Fall: Die Nennerpotenzreihe ist für $z = z_0$ gleich null. Ist ihr erster von null verschiedener Koeffizient b_m $(m > 0)$, so kommt man durch Ausklammern von $(z - z_0)^m$ aus der Nennerpotenzreihe zum ersten Fall zurück:

$$\left(\sum_{\nu=m}^{\infty} b_\nu(z - z_0)^\nu\right)^{-1} \sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu = (z - z_0)^{-m}\left(\sum_{\nu=0}^{\infty} b_{\nu+m}(z - z_0)^\nu\right)^{-1} \sum_{\nu=0}^{\infty} a_\nu(z - z_0)^\nu.$$

Also hat das Ergebnis die Struktur

$$(z - z_0)^{-m} \sum_{\nu=0}^{\infty} d_\nu(z - z_0)^\nu = d_0(z - z_0)^{-m} + d_1(z - z_0)^{-m+1} + \ldots + d_{m-1}(z - z_0)^{-1} + d_m + d_{m+1}(z - z_0) + d_{m+2}(z - z_0)^2 + \ldots$$

Man spricht von einer *Laurent-Reihe mit abbrechendem Hauptteil* (Hauptteil: $\sum_{\nu=1}^{m} d_{m-\nu}(z - z_0)^{-\nu}$).

Es liegt an der Stelle $z = z_0$ ein *Pol höchstens m-ter Ordnung* vor. Ist bekannt, daß $d_0 \neq 0$ ist, so ist es ein *Pol m-ter Ordnung*. Der Konvergenzradius der Ergebnisreihe ist mindestens gleich der Entfernung zwischen der Entwicklungsstelle z_0 und der von z_0 verschiedenen ihr jedoch am nächsten gelegenen Nullstelle der Nennerpotenzreihe.

Einsetzen einer Potenzreihe in eine andere. Aus $w = f(u) = \sum_{\nu=0}^{\infty} a_\nu(u - u_0)^\nu$ und $u = g(z) = \sum_{\mu=0}^{\infty} b_\mu(z - z_0)^\mu$ folgt unter der Voraussetzung $u_0 = b_0$ für die mittelbare Funktion $w = f(g(z)) = \sum_{\nu=0}^{\infty} a_\nu\left(\sum_{\mu=1}^{\infty} b_\mu(z - z_0)^\mu\right)^\nu = \sum_{k=0}^{\infty} c_k(z - z_0)^k$, wobei zunächst durch Potenzreihenmultiplikation die Potenzen $\left(\sum_{\mu=1}^{\infty} b_\mu(z - z_0)^\mu\right)^\nu$ in Potenzreihen entwickelt werden, um schließlich durch Ordnen nach Potenzen von $z - z_0$ die unbekannten Koeffizienten c_k zu erhalten.

Umkehrung von Potenzreihen. Die Gleichung $w = \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu$ (1) ist im Falle $a_1 \neq 0$ in einer gewissen Umgebung von $w = a_0$ eindeutig nach z auflösbar; die Auflösung $z = z(w)$ hat die Struktur $z = z_0 + \sum_{\mu=1}^{\infty} b_\mu (w - a_0)^\mu$ (2). Zur Bestimmung der b_μ setzt man (2) in (1) ein, entwickelt die entstehende rechte Seite von (1) nach der Methode „Einsetzen einer Potenzreihe in eine andere" und führt schließlich bezüglich der Potenzen von $w - a_0$ einen Koeffizientenvergleich durch.

Differentiation und Integration von Potenzreihen kann gliedweise erfolgen: $\frac{d}{dz} \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu = \sum_{\nu=1}^{\infty} \nu a_\nu (z - z_0)^{\nu-1}$

$$= \sum_{\nu=0}^{\infty} (\nu + 1) a_{\nu+1} (z - z_0)^\nu,$$

$$\int \sum_{\nu=0}^{\infty} a_\nu (z - z_0)^\nu dz = C + \sum_{\nu=0}^{\infty} a_\nu \frac{1}{\nu + 1} (z - z_0)^{\nu+1}.$$

Entwicklung im Unendlichen. Man sagt, die Funktion $w = f(z)$ werde an der Stelle $z = \infty$ in eine Laurent-Reihe entwickelt, wenn mit $t = 1/z$ die Funktion $g(t) = f(z)$ an der Stelle $t = 0$ in eine Laurent-Reihe entwickelt wird.

13.3.4.6. Angabe einiger elementarer Funktionen durch Reihen-Entwicklungen. Exponentialfunktion.

$$e^z = \sum_{\nu=0}^{\infty} \frac{z^\nu}{\nu!} = 1 + z + z^2/2! + \ldots, \ |z| < \infty.$$

Kosinusfunktion. $\cos z = (1/2)(e^{iz} + e^{-iz}) =$

$$\sum_{\nu=0}^{\infty} (-1)^\nu \frac{z^{2\nu}}{(2\nu)!} = 1 - z^2/2! + \ldots, \ |z| < \infty.$$

Sinusfunktion. $\sin z = (1/2i)(e^{iz} - e^{-iz}) =$

$$\sum_{\nu=0}^{\infty} (-1)^\nu \frac{z^{2\nu+1}}{(2\nu + 1)!} = z - \frac{z^3}{3!} + \ldots, \ |z| < \infty.$$

Tangensfunktion. $\tan z = \frac{\sin z}{\cos z} = \frac{1}{i} \frac{e^{2iz} - 1}{e^{2iz} + 1}$

$$= z + \frac{1}{3} z^3 + \frac{2}{15} z^5 + \ldots, \ |z| < \pi/2.$$

Kotangensfunktion.

$$\cot z = \frac{\cos z}{\sin z} = \frac{1}{z} - \frac{1}{3} z^3 - \frac{1}{45} z^3 - \ldots 0 < |z| < \pi.$$

Hyperbolische Kosinusfunktion.

$$\cosh z = \cos(iz) = (1/2)(e^z + e^{-z}) = \sum_{\nu=0}^{\infty} \frac{z^{2\nu}}{(2\nu)!}.$$

Hyperbolische Sinusfunktion.

$$\sinh z = -i \sin(iz) = (1/2)(e^z - e^{-z}) = \sum_{\nu=0}^{\infty} \frac{z^{2\nu+1}}{(2\nu + 1)!}.$$

Hyperbolische Tangensfunktion.

$$\tanh z = -i \tan(iz) = \frac{e^{2z} - 1}{e^{2z} + 1}.$$

Hyperbolische Kotangensfunktion. $\coth z = i \cot(iz)$.

Logarithmusfunktion. Die Gleichung $z = e^w$ hat bei gegebenem $z = |z| e^{i\varphi} (-\pi < \varphi \leqq \pi)$ (vgl. 13.1.3.3.) unendlich viele Lösungen $w = \ln|z| + i\varphi + 2k\pi i$ (1) ($k = 0, \pm 1, \pm 2, \ldots$) [ln = *natürlicher Logarithmus* (*Logarithmus zur Basis* e)]. Bei festem k stellt die rechte Seite von (1) den *k-ten Zweig des (komplexen) Logarithmus* dar. Der nullte Zweig heißt *Hauptwert des Logarithmus* und man schreibt

$\ln z = \ln|z| + i\varphi$ ($z = |z| e^{i\varphi}$, $-\pi < \varphi \leqq \pi$).

$$\ln z = \sum_{\nu=1}^{\infty} (-1)^{\nu+1} \nu^{-1} (z - 1)^\nu \ (0 < |z - 1| < 1).$$

Potenzfunktion. Die *allgemeine* Potenz umfaßt die Zweige $z^\alpha = e^{\alpha(\ln|z| + i\varphi + 2k\pi i)}$ ($z = |z| e^{i\varphi}$, $-\pi < \varphi \leqq \pi$, $k = 0, \pm 1, \ldots$). Der nullte Zweig ($k = 0$) heißt *Hauptwert der Potenz*. Im Fall $\alpha = n = 0, \pm 1, \ldots$ sind alle Zweige einander gleich. Im Fall $\alpha = p/q$ (p, q ganz, $q > 0, p, q$ teilerfremd) sind in $z^\alpha = \sqrt[q]{z^p}$ genau q Zweige zusammengefaßt. Ist z reell und darüber hinaus größer als null, so versteht man unter z^α den Hauptwert.

Bsp. 13.3.4–3. *Fakultätenfunktion* (*Gammafunktion*):

$$z! = \Gamma(z + 1) = \int_0^\infty e^{-t} t^z dt \quad (\operatorname{Re} z > -1).$$

Binomische Entwicklung. Hauptwert von $(1 + z)^\alpha = \sum_{\nu=0}^{\infty} \binom{\alpha}{\nu} z^\nu$ ($|z| < \infty$, falls $\alpha = 0, 1, 2, \ldots$; sonst $|z| < 1$).

Binomialkoeffizient:

$$\binom{\alpha}{0} = 1, \ \binom{\alpha}{\nu} = \frac{\alpha(\alpha - 1) \cdot \ldots \cdot (\alpha - \nu + 1)}{1 \cdot 2 \cdot \ldots \cdot \nu} \quad (\nu = 1, 2, \ldots).$$

Arkussinusfunktion. $z = \sin w \Leftrightarrow w = \arcsin z + 2k\pi$ oder $w = -\arcsin z + (2k + 1)\pi$ ($k = 0, \pm 1, \ldots$)

[$\Leftrightarrow$ bedeutet: „gleichbedeutend mit", „äquivalent zu"].

$$\arcsin z = \sum_{\nu=0}^{\infty} \frac{1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2\nu - 1)}{2 \cdot 4 \cdot 6 \cdot \ldots \cdot (2\nu)(2\nu + 1)} z^{2\nu+1} \quad (|z| < 1).$$

Arkuskosinusfunktion. $z = \cos w \Leftrightarrow w = \pm \arccos z + 2k\pi$ ($k = 0, \pm 1, \ldots$). $\arccos z = \pi/2 - \arcsin z$.

Arkustangensfunktion. $z = \tan w \Leftrightarrow w = \arctan z + k\pi$ ($k = 0, \pm 1, \ldots$). $\arctan z = \sum_{\nu=0}^{\infty} (-1)^\nu \frac{1}{2\nu + 1} z^{2\nu+1}$, ($|z| < 1$).

$\arctan z = C + \sum_{\nu=0}^{\infty} (-1)^{\nu+1} \frac{1}{2\nu + 1} z^{-(2\nu+1)}$ (z ist reell mit $|z| > 1$; $C = \pi/2$, falls $z > 0$; $C = -\pi/2$, falls $z < 0$).

Arkuskotangensfunktion. $z = \cot w \Leftrightarrow w = \operatorname{arccot} z + k\pi$ ($k = 0, \pm 1, \ldots$). $\operatorname{arccot} z = \pi/2 - \arctan z$.

Areakosinusfunktion. $z = \cosh w \Leftrightarrow w = \pm \operatorname{arccosh} z + 2k\pi i$ ($k = 0, \pm 1, \ldots$).

$\operatorname{arccosh} z = \ln(2z) - \sum_{\nu=1}^{\infty} \frac{1 \cdot 3 \cdot \ldots \cdot (2\nu - 1)}{2 \cdot 4 \cdot \ldots \cdot (2\nu)} \frac{1}{2\nu} z^{-2\nu}$ (z reell, $|z| > 1$).

Areasinusfunktion. $z = \sinh w \Leftrightarrow w = \operatorname{arsinh} z + 2k\pi i$ oder $w = -\operatorname{arsinh} z + (2k + 1)\pi i$.

$\operatorname{arsinh} z = \sum_{\nu=0}^{\infty} (-1)^\nu \frac{1 \cdot 3 \cdot \ldots \cdot (2\nu - 1)}{2 \cdot 4 \cdot \ldots \cdot (2\nu)} \frac{1}{2\nu + 1} z^{2\nu+1}$ ($|z| < 1$).

Areatangensfunktion. $z = \tanh w \Leftrightarrow w = \operatorname{artanh} z + k\pi i$ ($k = 0, \pm 1, \ldots$). $\operatorname{artanh} z = \sum_{\nu=0}^{\infty} \frac{1}{2\nu + 1} z^{2\nu+1}$ ($|z| < 1$).

Areakotangensfunktion. $z = \coth w \Leftrightarrow w = \operatorname{arcoth} z + k\pi i$ ($k = 0, \pm 1, \ldots$). $\operatorname{arcoth} z = \sum_{\nu=0}^{\infty} \frac{1}{2\nu + 1} z^{-(2\nu+1)}$ ($|z| > 1$).

13.3.4.7. Taylor-Entwicklung von Funktionen, die von mehreren unabhängigen Variablen abhängen. Die Funktionen aus 13.3.3.3. werden fortgesetzt zu *komplexen* Funktionen, die von *komplexen* unabhängigen Variablen abhängen; die Bezeichnung aus 13.3.3.3. wird jedoch beibehalten. Ist die Funktion $U = U(x, y, z)$ an der Stelle (x_0, y_0, z_0) sowohl bezüglich x als auch y und z holomorph, so strebt in der dazugehörigen Taylor-Formel (vgl. 13.3.4.3.) das Restglied für $n \to \infty$ nach null, falls (x, y, z) sich in einer gewissen Umgebung von (x_0, y_0, z_0) aufhält. Somit ergibt sich dort

die Taylor-Entwicklung

$$U(x,y,z) = U[x_0 + (x - x_0), y_0 + (y - y_0), z_0 + (z - z_0)]$$
$$= \sum_{\nu=0}^{\infty} \frac{1}{\nu!} \left\{ \left[(x - x_0) \frac{\partial}{\partial \bar{x}} + (y - y_0) \frac{\partial}{\partial \bar{y}} + (z - z_0) \frac{\partial}{\partial \bar{z}} \right]^\nu U(\bar{x}, \bar{y}, \bar{z}) \right\}_{(\bar{x},\bar{y},\bar{z}) = (x_0, y_0, z_0)}$$
$$= U + \frac{\partial U}{\partial x}(x - x_0) + \frac{\partial U}{\partial y}(y - y_0) + \frac{\partial U}{\partial z}(z - z_0)$$
$$+ \frac{1}{2} \left\{ \frac{\partial^2 U}{\partial x^2}(x - x_0)^2 + \frac{\partial^2 U}{\partial y^2}(y - y_0)^2 + \frac{\partial^2 U}{\partial z^2}(z - z_0)^2 \right.$$
$$+ 2\frac{\partial^2 U}{\partial x \partial y}(x - x_0)(y - y_0) + 2\frac{\partial^2 U}{\partial x \partial z}(x - x_0)(z - z_0)$$
$$\left. + 2\frac{\partial^2 U}{\partial y \partial z}(y - y_0)(z - z_0) \right\} + \frac{1}{3!}\{\ldots\} + \ldots,$$

wobei hinter dem letzten Gleichheitszeichen für die Funktion U und ihre sämtlichen partiellen Ableitungen die jeweiligen Werte an der Stelle (x_0, y_0, z_0) einzusetzen sind. Es wurde beim ausführlichen Aufschreiben der Entwicklung noch der *Satz von Schwarz* benutzt, der aussagt, daß es bei Bildung von höheren partiellen Ableitungen auf die Reihenfolge der Differentiationen nicht ankommt. Analoge Entwicklungen gelten für Funktionen, die von zwei, vier, fünf, ... unabhängigen Variablen abhängen.

13.3.5. Kettenregel

Die mittelbare Funktion F mit $F(u) = F_1[F_2(u)]$ ist an der Stelle u_0 Fréchet-differenzierbar (vgl. 13.3.3.), falls F_2 an der Stelle u_0 und F_1 an der Stelle $F_2(u_0)$ Fréchet-differenzierbar sind. Mit den Abkürzungen $T = F'(u_0)$, $T_1 = F_1'[F_2(u_0)]$, $T_2 = F_2'(u_0)$ lautet die *Kettenregel* $T = T_1 T_2$, d.h. $Th = T_1(T_2 h)$.

Bsp. 13.3.5–1. Mit $U = U(y_1, y_2, y_3)$, $y_k = y_k(x_1, x_2)$ $(k = 1, 2, 3)$ werde die mittelbare Funktion $f(x_1, x_2) = U(y_1(x_1, x_2), y_2(x_1, x_2), y_3(x_1, x_2))$ gebildet. Die Fréchet-Ableitung von $U = U(y_1, y_2, y_3)$ kann durch die einspaltige Matrix $(\partial U/\partial y_k)$ $(k = 1, 2, 3)$ (vgl. 13.3.3.3.) angegeben werden. Die Fréchet-Ableitung der einspaltigen Matrix $(y_k(x_1, x_2))$ $(k = 1, 2, 3)$ ist durch die Matrix $(\partial y_k/\partial x_l)$ $[k = 1, 2, 3;\ l = 1, 2;$ Typ $(3,2)]$ angebbar. Der Hintereinanderausführung $T_1 T_2$ der stetigen linearen Operatoren T_1 und T_2 entspricht hier die Matrizenmultiplikation $(\partial U/\partial y_k)^T (\partial y_k/\partial x_l)$. Also ist die Fréchet-Ableitung der obigen mittelbaren Funktion f gegeben durch die einspaltige Matrix $(\partial f/\partial x_l)$ $(l = 1, 2)$ mit $\frac{\partial f}{\partial x_l} = \frac{\partial U}{\partial y_k} \frac{\partial y_k}{\partial x_l}$ (Einsteinsche Summationsverabredung!).

13.3.5.1. Kurven, die auf einer gegebenen Fläche liegen. Setzt man in der Parameterdarstellung einer Fläche A: $\boldsymbol{r}(u, v) = x_k(u, v)\boldsymbol{e}_k$ (vgl. 13.3.1.3.) für $u = u(t), v = v(t)$ ein, so erhält man die Parameterdarstellung einer Kurve C, die auf A liegt: $\boldsymbol{r}[u(t), v(t)] = x_k[u(t), v(t)]\boldsymbol{e}_k$ (vgl. 13.3.1.3.). Ein Tangentenvektor an die Kurve C (vgl. 13.3.3.4.) ergibt sich mittels der Kettenregel zu $d\boldsymbol{r}/dt = (\partial \boldsymbol{r}/\partial u)\dot{u} + (\partial \boldsymbol{r}/\partial v)\dot{v} = [(\partial x_k/\partial u)\dot{u} + (\partial x_k/\partial v)\dot{v}]\boldsymbol{e}_k$.

13.3.5.2. Kurvendarstellung in krummlinigen Koordinaten. Werden krummlinige Koordinaten (u_1, u_2, u_3) durch $\boldsymbol{r}(u_1, u_2, u_3) = x_k(u_1, u_2, u_3)\boldsymbol{e}_k$ (vgl. 13.3.1.3.) eingeführt, so ist $\boldsymbol{r}[u_1(t), u_2(t), u_3(t)]$ die Parameterdarstellung einer Kurve. Ein Tangentenvektor von C ist $d\boldsymbol{r}/dt = (\partial \boldsymbol{r}/\partial u_l)\dot{u}_l = (\partial x_k/\partial u_l)\dot{u}_l \boldsymbol{e}_k$ (Einsteinsche Summationsverabredung!).

13.3.5.3. Zusammenhang zwischen Kurvenparameter und Bogenlängenkoordinate. Arbeitet man neben dem Kurvenparameter t mit dem Bogenlängenparameter s der Kurve, so gilt $d\boldsymbol{r}/dt = (d\boldsymbol{r}/ds)(ds/dt)$. Hierbei ist $d\boldsymbol{r}/ds$ ein Tangentenvektor der Kurve mit dem absoluten Betrag Eins. Ist t die Zeit, so ist $d\boldsymbol{r}/dt$ die vektorielle und ds/dt die skalare Geschwindigkeit.

13.3.5.4. Differentiation implizit dargestellter Funktionen. Gegeben seien die m Funktionen $F_k(x_1, \ldots, x_n, y_1, \ldots, y_m)$ $(k = 1, \ldots, m)$. Ferner sei $x_{10}, \ldots, x_{n0}, y_{10}, \ldots, y_{m0})$ eine Lösung des Gleichungssystems $F_k = 0$ $(k = 1, \ldots, m)$. Ist die Determinante der Matrix $(\partial F_k/\partial y_l)$ $(k = 1, \ldots, m;\ l = 1, \ldots, m)$ ungleich Null, so ist das Gleichungssystem $F_k(x_1, \ldots, x_n, y_1, \ldots, y_m) = 0$ $(k = 1, \ldots, m)$ in einer Umgebung von $x_{10}, \ldots, x_{n0}, y_{10}, \ldots, y_{m0})$ in eindeutiger Weise nach $y_1, \ldots, y_m$ auflösbar: $y_k = y_k(x_1, \ldots, x_n)$ $(k = 1, \ldots, m)$. Setzt man diese Funktionen in $F_k = 0$ ein, so entstehen die Identitäten $F_k[x_1, \ldots, x_n, y_1(x_1, \ldots, x_n), \ldots, y_m(x_1, \ldots, x_n)] = 0$. Differenziert man diese jeweils partiell nach x_l, so ergibt sich gemäß der Kettenregel $\frac{\partial F_k}{\partial x_l} + \sum_{p=1}^{m} \frac{\partial F_k}{\partial y_p}\frac{\partial y_p}{\partial x_l} = 0$, also ein lineares Gleichungssystem für die partiellen Ableitungen $\partial y_p/\partial x_l$ $(p = 1, \ldots, m;\ \ l = 1, \ldots, n;\ \ l = \text{fest})$, welches wegen $\det(\partial F_k/\partial y_p) \neq 0$ genau eine Lösung besitzt (vgl. 13.2.3.2.).

Bsp. 13.3.5–1. Ist (x_0, y_0) ein Punkt des Kreises $x^2 + y^2 - R^2 = 0$, so ist im Fall $y_0 \neq 0$ diese Gleichung in einer Umgebung von (x_0, y_0) eindeutig nach y auflösbar – im vorliegenden Fall ist die Auflösung (ausnahmsweise) durch elementare Funktionen angebbar, nämlich $y = \sqrt{R^2 - x^2}$, falls $y_0 > 0$ und $y = -\sqrt{R^2 - x^2}$, falls $y_0 < 0$. Das Einsetzen der Lösung $y = y(x)$ in die Kreisgleichung führt zur Identität $x^2 + [y(x)]^2 - R^2 \equiv 0$. Durch Differentiation nach x ergibt sich $2x + 2yy' = 0$ und damit $y' = -x/y$.

13.3.6. Produktregel

Mit dem gegebenen stetigen bilinearen Operator B (vgl. 13.3.4.1.) werde die mittelbare Funktion F mit $F(u) = B[F_1(u), F_2(u)]$ gebildet. Dann ist F an der Stelle u_0 Fréchet-differenzierbar, falls F_1 und F_2 dort Fréchet-differenzierbar sind, und es gilt mit den Abkürzungen $T = F'(u_0)$, $T_1 = F_1'(u_0)$, $T_2 = T_2'(u_0)$ für das Fréchet-Differential Th von F die *Produktregel* $Th = B[T_1 h, F_2(u_0)] + B[F_1(u_0), T_2 h]$.

13.3.6.1. Frenetsche Ableitungsformeln. Der Tangentenvektor $\boldsymbol{t} = d\boldsymbol{r}/ds$ $(s = \text{Bogenlängenkoordinate})$ an eine Raumkurve C hat den absoluten Betrag 1 (vgl. 13.3.5.3.), also gilt die Identität $\boldsymbol{t} \cdot \boldsymbol{t} \equiv 1$. Differentiation nach s führt gemäß der Produktregel zu $2\boldsymbol{t} \cdot d\boldsymbol{t}/ds = 0$. Also steht der Vektor $d\boldsymbol{t}/ds$ auf dem Vektor $\boldsymbol{t}$ senkrecht. Der absolute Betrag von $\boldsymbol{t}' = d\boldsymbol{t}/ds$ heißt *Krümmung* $\varkappa$ der Kurve C ($\varrho = (1/\varkappa) = \textit{Krümmungsradius}$). Also gilt $\boldsymbol{t}' = \varkappa \boldsymbol{n}$ $(|\boldsymbol{n}| = 1)$ mit dem *Normaleneinheitsvektor* $\boldsymbol{n}$ der Kurve C. Die Vektoren $\boldsymbol{t}, \boldsymbol{n}$ und $\boldsymbol{b} = \boldsymbol{t} \times \boldsymbol{n}$ ($\boldsymbol{b}$ = *Binormalenvektor*) bilden ein Orthonormalsystem (*orthonormale Basis*, vgl. 13.1.4.1.), das als *begleitendes Dreibein* der Kurve C bezeichnet wird. Es gelten die *Frenetschen Ableitungsformeln* (Striche bedeuten Ableitung nach der Bogenlängenkoordinate s): $\boldsymbol{t}' = \varkappa \boldsymbol{n}$, $\boldsymbol{n}' = -\varkappa \boldsymbol{t} + \tau \boldsymbol{b}$, $\boldsymbol{b}' = -\tau \boldsymbol{n}$ mit der Windung $\tau = -\boldsymbol{b}' \cdot \boldsymbol{n} = -(\boldsymbol{t} \times \boldsymbol{n})' \cdot \boldsymbol{n}$. Bei Beachtung von $\boldsymbol{n} = \varrho \boldsymbol{t}'$ und nach Anwendung der Produkt- und Kettenregel gilt $\tau = \varrho^2(\boldsymbol{t}\boldsymbol{t}'\boldsymbol{t}'')$ (vgl. 13.2.5.1., Spatprodukt).

Hauptsatz der Kurventheorie. Gibt man stetige Funktionen $\varkappa(s) > 0$ und $\tau(s)$ vor, dann existiert, abgesehen von der

räumlichen Lage, genau eine Kurve, die s als Bogenlängenkoordinate besitzt und deren Krümmung bzw. Windung mit $\varkappa(s)$ bzw. $\tau(s)$ übereinstimmt.

Bsp. 13.3.6–1. Eine Punktmasse bewege sich längs der Kurve $\boldsymbol{r} = \boldsymbol{r}(t)$ (t = Zeit). Die *vektorielle Geschwindigkeit* ist $\boldsymbol{v} = \dot{\boldsymbol{r}}(t) = \mathrm{d}\boldsymbol{r}/\mathrm{d}t = (\mathrm{d}\boldsymbol{r}/\mathrm{d}s)(\mathrm{d}s/\mathrm{d}t) = \boldsymbol{t}v$ ($v = |\boldsymbol{v}| = \mathrm{d}s/\mathrm{d}t$ = *skalare Geschwindigkeit*). Die *vektorielle Beschleunigung* ergibt sich zu $\dot{\boldsymbol{v}}(t) = (\mathrm{d}/\mathrm{d}t)(v\boldsymbol{t}) = (\mathrm{d}v/\mathrm{d}t)\boldsymbol{t} + v(\mathrm{d}\boldsymbol{t}/\mathrm{d}t)$ mit $\mathrm{d}\boldsymbol{t}/\mathrm{d}t = (\mathrm{d}\boldsymbol{t}/\mathrm{d}s)(\mathrm{d}s/\mathrm{d}t) = \boldsymbol{t}'v = \varkappa\boldsymbol{n}v$; also ist $\dot{\boldsymbol{v}}(t) = (\mathrm{d}v/\mathrm{d}t)\boldsymbol{t} + (v^2/\varrho)\boldsymbol{n}$, in Worten: Die vektorielle Beschleunigung ist gleich der Summe aus der (*vektoriellen*) *Tangentialbeschleunigung* und der (vektoriellen) *Zentripetalbeschleunigung*.

13.3.6.2. Vektoranalysis.

Nabla-Operator. $\nabla = \partial/\partial\boldsymbol{r} = \boldsymbol{e}_k\partial/\partial x_k$.

Gradient. $\mathbf{grad}\, U = \nabla U$ = Fréchet-Ableitung von U $= (\partial U/\partial x_k)\boldsymbol{e}_k$ (vgl. 13.2.6.7., 13.3.3.4.).

Gradient eines Vektors $\boldsymbol{v}$ heißt der Tensor zweiter Stufe mit den Koordinaten $\partial v_k/\partial x_l$ (vgl. 13.3.3.7.).

Divergenz. $\operatorname{div}\boldsymbol{v} = \nabla\cdot\boldsymbol{v}$ = Verjüngung des Tensors $\partial v_k/\partial x_l$ $= \partial v_k/\partial x_k$ (vgl. 13.2.6.7., 13.3.3.7.).

Rotation. $\mathbf{rot}\,\boldsymbol{v} = \nabla\times\boldsymbol{v} = -2$ mal (demjenigen Vektor, mit dessen Hilfe der antimetrische Anteil des Tensors $\partial v_k/\partial x_l$ bestimmt werden kann) $= -2\cdot(1/2)\varepsilon_{klm}(\partial v_k/\partial x_l)\boldsymbol{e}_m$ $= \varepsilon_{lkm}(\partial/\partial x_l)v_k\boldsymbol{e}_m$ (vgl. 13.2.6.7.; 13.3.3.7.; 13.2.6.2., Zerlegung eines Tensors zweiter Stufe in seinen symmetrischen und antimetrischen Anteil, Bsp. 13.2.6–10).

Bilinearer Operator $U(x_1,x_2,x_3)\,V(x_1,x_2,x_3)$. Gemäß der Produktregel ergibt sich für sein Fréchet-Differential $(\partial U/\partial x_l)h_lV + U(\partial V/\partial x_l)h_l$ $(\boldsymbol{h} = h_l\boldsymbol{e}_l)$. Hieraus folgt $\mathbf{grad}\,(UV) = \nabla(UV) = V\nabla U + U\nabla V = V\mathbf{grad}\,U + U\mathbf{grad}\,V$.

Bilinearer Operator $U(x_1,x_2,x_3)\,\boldsymbol{v}(x_1,x_2,x_3)$. Gemäß der Produktregel ergibt sich für sein Fréchet-Differential $(\partial U/\partial x_l)h_lv_k\boldsymbol{e}_k + U(\partial v_k/\partial x_l)h_l\boldsymbol{e}_k$ $(\boldsymbol{v} = v_k\boldsymbol{e}_k)$. Hieraus folgt einerseits $\operatorname{div}(U\boldsymbol{v}) = \nabla\cdot(U\boldsymbol{v}) = \boldsymbol{v}\cdot\nabla U + U\nabla\cdot\boldsymbol{v} = \boldsymbol{v}\cdot\mathbf{grad}\,U + U\operatorname{div}\boldsymbol{v}$ und andererseits $\mathbf{rot}(U\boldsymbol{v}) = \nabla\times(U\boldsymbol{v}) = U\nabla\times\boldsymbol{v} + (\nabla U)\times\boldsymbol{v} = U\mathbf{rot}\,\boldsymbol{v} + \mathbf{grad}\,U\times\boldsymbol{v}$.

Bilinearer Operator $\boldsymbol{v}(x_1,x_2,x_3)\cdot\boldsymbol{w}(x_1,x_2,x_3)$. Gemäß der Produktregel ergibt sich für sein Fréchet-Differential $(\partial v_k/\partial x_l)h_lw_k + v_k(\partial w_k/\partial x_l)h_l$. Hieraus folgt

$$\mathbf{grad}\,(\boldsymbol{v}\cdot\boldsymbol{w}) = \boldsymbol{v}\times\mathbf{rot}\,\boldsymbol{w} + \boldsymbol{w}\times\mathbf{rot}\,\boldsymbol{v} + (\boldsymbol{v}\cdot\nabla)\boldsymbol{w} + (\boldsymbol{w}\cdot\nabla)\boldsymbol{v}.$$

$$\left[(\boldsymbol{u}\cdot\nabla)\boldsymbol{v} = \left(u_k\frac{\partial}{\partial x_k}\right)v_l\boldsymbol{e}_l = u_k(\partial v_l/\partial x_k)\boldsymbol{e}_l\right]$$

Bilinearer Operator $\boldsymbol{v}(x_1,x_2,x_3)\times\boldsymbol{w}(x_1,x_2,x_3)$. Gemäß der Produktregel ergibt sich für sein Fréchet-Differential $\varepsilon_{kpq}\big((\partial v_k/\partial x_l)h_lw_p + v_k(\partial w_p/\partial x_l)h_l\big)\boldsymbol{e}_q$. Hieraus folgt einerseits $\operatorname{div}(\boldsymbol{v}\times\boldsymbol{w}) = \nabla\cdot(\boldsymbol{v}\times\boldsymbol{w}) = \boldsymbol{w}\cdot(\nabla\times\boldsymbol{v}) - \boldsymbol{v}\cdot(\nabla\times\boldsymbol{w})$ $= \boldsymbol{w}\cdot\mathbf{rot}\,\boldsymbol{v} - \boldsymbol{v}\cdot\mathbf{rot}\,\boldsymbol{w}$ und andererseits $\mathbf{rot}(\boldsymbol{v}\times\boldsymbol{w}) = \boldsymbol{v}\operatorname{div}\boldsymbol{w} - \boldsymbol{w}\operatorname{div}\boldsymbol{v} + (\boldsymbol{w}\cdot\nabla)\boldsymbol{v} - (\boldsymbol{v}\cdot\nabla)\boldsymbol{w}$.

Zweifache Anwendung des Nablaoperators. $\operatorname{div}\mathbf{grad}\,U = \nabla\cdot\nabla U = \Delta U$ (vgl. 13.2.6.7.). $\operatorname{div}(\mathbf{rot}\,\boldsymbol{v}) = \nabla\cdot(\nabla\times\boldsymbol{v}) = 0$, $\mathbf{rot}(\mathbf{grad}\,U) = \nabla\times(\nabla U) = 0$, $\mathbf{rot}(\mathbf{rot}\,\boldsymbol{v}) = \nabla\times(\nabla\times\boldsymbol{v})$ $= \mathbf{grad}(\operatorname{div}\boldsymbol{v}) - \Delta\boldsymbol{v}$ mit $\Delta\boldsymbol{v} = (\Delta v_k)\boldsymbol{e}_k$ (vgl. 13.2.6.7.).

13.4. Integralrechnung

13.4.1. Integrierbarkeit

13.4.1.1. Maß. Ein spezielles Maß ist das in 13.1.6.6. genannte Wahrscheinlichkeitsmaß P eines Wahrscheinlichkeitsraumes $[\Omega, B, P]$. Ein (allgemeines) *Maß* μ entsteht definitionsgemäß aus P, wenn man beim Übergang vom P zum μ zwar das Axiom der σ-Additivität beibehält, jedoch die Axiome $0 \leqq P(E) \leqq 1$, $P(\Omega) = 1$ durch die geringere Forderung $0 \leqq \mu(E) = \infty$ ($E\in B$, ∞ zugelassen!), $\mu(\emptyset) = 0$ ($\emptyset$: leere Menge) ersetzt. Es entsteht der *Maßraum* $[\Omega, B, \mu]$.

13.4.1.2. Integral der einfachen Funktionen. Eine reelle Funktion g, deren Definitionsbereich die Menge Ω eines Maßraumes $[\Omega, B, \mu]$ ist, heißt *einfach*, wenn sie nur endlich viele Werte $\alpha_1, \ldots, \alpha_n$ annimmt, wobei diejenigen Teilmengen Ω_ν von Ω, auf denen g jeweils einen Wert $\alpha_\nu \neq 0$ hat, einerseits zu B gehören und andererseits ein Maß $\mu(\Omega_\nu)$ haben, das nicht gleich unendlich ist. Dann heißt $\int_\Omega g\,\mathrm{d}\mu = \sum_{\nu=1}^{n}\alpha_\nu\mu(\Omega_\nu)$ das *Integral der einfachen Funktion* g bezüglich μ.

13.4.1.3. Definition des Integrals. Gegeben sei eine reelle Funktion g, deren Definitionsbereich gleich der Menge Ω eines Maßraumes $[\Omega, B, \mu]$ ist. Existiert eine Folge einfacher Funktionen g_n ($n = 1, 2, \ldots$) derart, daß einerseits $\int_\Omega |g_n - g_m|\,\mathrm{d}\mu$ beliebig klein wird, falls m und n hinreichend groß sind, und andererseits $\lim_{n\to\infty}\mu(E_n) = 0$ für alle $\sigma > 0$ gilt, wobei E_n diejenige – voraussetzungsgemäß zu B gehörende – Teilmenge von Ω ist, für deren Elemente ω die Ungleichheit $|g_n(\omega) - g(\omega)| \geqq \sigma$ gilt, so heißt g eine *μ-integrierbare* (auch: *μ-summierbare*) Funktion und $\lim_{n\to\infty}\int_\Omega g_n\,\mathrm{d}\mu$ das (*bestimmte*) *Integral* von g bezüglich μ. Es wird mit $\int_\Omega g\,\mathrm{d}\mu$, $\int_\Omega g(\omega)\,\mathrm{d}\mu(\omega)$ oder $\int_\Omega g(\omega)\,\mu(\mathrm{d}\omega)$ bezeichnet. Für eine Teilmenge Ω' von Ω mit $\Omega' \in B$ definiert man $\int_{\Omega'} g(\omega)\,\mu(\mathrm{d}\omega) = \int_\Omega \chi_{\Omega'}(\omega)\,g(\omega)\,\mu(\mathrm{d}\omega)$, falls $\chi_{\Omega'}\cdot g$ μ-integrierbar ist, wobei $\chi_{\Omega'}$ *Indikatorfunktion* von Ω' heißt, d.h. es gilt $\chi_{\Omega'}(\omega) = 1$ für $\omega\in\Omega'$ und $\chi_{\Omega'}(\omega) = 0$ für $\omega\notin\Omega'$.

13.4.2. Integration von reellen Funktionen einer reellen unabhängigen Variablen

13.4.2.1. Lebesgue-Stieltjes-Maße. Es sei F eine reelle Funktion, deren Definitionsbereich die Menge $\Omega = R^1$ aller reellen Zahlen ist, und einerseits monoton nicht-fallend (es gilt also $F(x_1) \leqq F(x_2)$ für $x_1 < x_2$), andererseits linksseitig stetig ist. Es gilt also für alle reellen Zahlen x_0 $\lim_{x\to x_0-0} F(x) = F(x_0)$, d. h. $\lim_{x\to x_0} F(x) = F(x_0)$ mit $x < x_0$ (vgl. 13.3.3.). Dann gibt es ein möglichst umfangreiches System B_F, das aus Teilmengen von R^1, den *bezüglich F Lebesgue-Stieltjes-meßbaren Mengen*, besteht und Definitionsbereich des *bezüglich F Lebesgue-Stieltjesschen Maßes* μ_F ist, wobei für jedes der Intervalle $(a, b) = \{x \mid a < x < b\}$ (lies: Menge aller x mit der Eigenschaft $a < x < b$) bzw. $[a, b] = \{x \mid a \leqq x \leqq b\}$ bzw. $(a, b] = \{x \mid a < x \leqq b\}$ bzw. $[a, b) = \{x \mid a \leqq x < b\}$ gilt: $\mu_F((a, b)) = F(b) - F(a+0)$ $\left(F(a+0) = \lim_{x\to a+0} F(x), \text{ d.h.} \lim_{x\to a} F(x) \text{ mit } x > a\right)$ bzw. $\mu_F([a, b]) = F(b+0) - F(a)$ bzw. $\mu_F((a, b]) = F(b+0) - F(a+0)$ bzw. $\mu_F([a, b)) = F(b) - F(a)$. Obwohl B_F umfangreich ist, gibt es – in der Praxis wenig interessierende – Teilmengen

von $\Omega = R^1$, die nicht meßbar sind. Ist speziell $F(x) = F_0(x) = x$, so heißen die zu $B_{F_0} = B$ gehörenden Mengen *Lebesgue-meßbare Mengen* und $\mu_{F_0} = \mu$ das *Lebesguesche Maß* des R^1. Insbesondere gilt $\mu((a, b)) = \mu([a, b]) = \mu((a, b]) = \mu([a, b)) = b - a$ und $\mu(R^1) = \infty$.

Bsp. 13.4.2–1. Eine Funktion F mit den oben geforderten Eigenschaften ist die Verteilungsfunktion einer Zufallsgröße (vgl. 13.3.1.7.), wobei darüber hinaus $\lim\limits_{x\to-\infty} F(x) = 0$, $\lim\limits_{x\to\infty} F(x) = 1$ und damit $\mu_F(R^1) = 1$ gilt.

13.4.2.2. Lebesgue-Stieltjes-Integral. Die Definition des *Lebesgue-Stieltjesschen Integrals* ist mittels 13.4.2.1. dem Abschnitt 13.4.1.3. zu entnehmen. Liegt das Integrationsintervall $a \leqq x \leqq b$ vor, so schreibt man für $\int\limits_{\Omega'} g(x)\,d\mu_F(x)$ kurz $\int\limits_a^b g(x)\,dF(x)$. Im Spezialfall $F(x) = F_0(x) = x$ schreibt man für $\int\limits_a^b g(x)\,dF_0(x)$ kurz $\int\limits_a^b g(x)\,dx$ und erhält das *Lebesgue-Integral*. Es stimmt mit dem *Riemann-Integral* überein, falls $g(x)$ beispielsweise auf $a \leqq x \leqq b$ stetig oder dort beschränkt und monoton oder dort beschränkt ist und eine endliche Anzahl von Unstetigkeitsstellen besitzt. Es gibt Integrale, die zwar im Lebesgueschen Sinn, jedoch nicht im Riemannschen Sinn existieren. Sie werden beispielsweise bei der Vervollständigung des Prä-Hilbert-Raumes der stetigen Funktionen zum Hilbert-Raum benötigt (vgl. Bsp. 13.1.4–9).

Bsp. 13.4.2–2. Ist F eine *Treppenfunktion* – d.h. die Menge der Funktionswerte von F ist endlich oder abzählbar-unendlich – so ist $\int\limits_a^b g(x)\,dF(x) = \sum\limits_\nu g(x_\nu)h_\nu$, wobei x_ν eine Sprungstelle von F und $h_\nu = F(x_\nu + 0) - F(x_\nu)$ den Sprung von F in x_ν bezeichnen.

Bsp. 13.4.2.–3. Ist $F(x)$ *absolutstetig* – d.h. es existiert die Ableitung $F'(x)$ (mit Ausnahme einer x-Menge vom Lebesguesschen Maß null) – dann ist $\int\limits_a^b g(x)\,dF(x) = \int\limits_a^b g(x)F'(x)\,dx$, also gleich dem Integral von $g(x)F'(x)$ bezüglich des Lebesgue-Maßes auf der Geraden.

Jede Funktion F, die gemäß 13.4.2.1. zu einem Maß μ_F Anlaß gibt, ist als Linearkombination einer Treppenfunktion (vgl. Bsp. 13.4.2–2), einer absolutstetigen Funktion (vgl. Bsp. 13.4.2–3) und einer – in der Praxis kaum vorkommenden – *singulären Funktion* F_S darstellbar.

Bsp. 13.4.2–4. Ist X eine beliebige Zufallsgröße (vgl. 13.3.1.7.) mit der Verteilungsfunktion $F(x)$, so heißt $\alpha_k = \int\limits_{-\infty}^{\infty} x^k dF(x)$ bzw. $\mu_k = \int\limits_{-\infty}^{\infty} (x - \alpha_1)^k dF(x)$ das *k-te (Anfangs-)Moment* bzw. *k-te zentrale Moment* von X $(k = 0, 1, 2, \ldots)$; speziell $\alpha_1 = EX$ *Erwartungswert* von X und $\mu_2 = D^2X$ die *Streuung* (auch: *Dispersion*, *mittlere quadratische Abweichung*, *Varianz*) von X. Es gilt:

$$\mu_k = \sum_{\nu=0}^{k} (-1)^{k-\nu}\binom{k}{\nu}\alpha_\nu\alpha_1^{k-\nu},$$

wobei $\alpha_k = i^{-k}\left.\dfrac{d^k\varphi(t)}{dt^k}\right|_{t=0}$

ist mit der *charakteristischen Funktion* $\varphi(t) = \varphi_X(t) = \int\limits_{-\infty}^{\infty} e^{itx}dF(x)$ (vgl. 13.4.7.1.). Es ist also speziell $EX = -i\varphi'_X(0)$ und $D^2X = -\varphi''_X(0) - (EX)^2$.

13.4.2.3. Integrationsregeln. Falls die links stehenden Lebesgue-Stieltjes-Integrale existieren, so gibt es auch die anderen und es gilt: *Linearität bezüglich des Integranden*:

$$c_1\int_a^b g_1(x)\,dF(x) + c_2\int_a^b g_2(x)\,dF(x) = \int_a^b (c_1g_1(x) + c_2g_2(x))\,dF(x).$$

Additivität bezüglich des Integrationsintervalls:

$$\int_a^c g(x)\,dF(x) + \int_c^b g(x)\,dF(x) = \int_a^b g(x)\,dF(x).$$

Partielle Integration: (($g(x)$ erfülle die Voraussetzungen von $F(x)$))

$$\int_a^b g(x)\,dF(x) + \int_a^b F(x)\,dg(x) = g(b)F(b) - g(a)F(a);$$

im Spezialfall $g(x) \equiv 1$ und differenzierbarem $F(x)$ ergibt sich der *Hauptsatz der Differential- und Integralrechnung*: $\int\limits_a^b f(x)\,dx = F(b) - F(a)$, wobei $F'(x) = f(x)$ gilt, d.h. $F(x)$ ist eine *Stammfunktion* von $f(x)$. Es ist $F(x)$ als Integral über f mit der oberen variablen Grenze x darstellbar, wobei für die Integrationsvariable ein anderer Buchstabe benutzt wird, weil die Bezeichnung x bereits für die obere Grenze verbraucht ist: $F(x) = \int\limits_a^x f(t)\,dt$. Alle Lösungen $F(x)$ von $F'(x) = f(x)$ unterscheiden sich nur um eine additive Konstante. Man schreibt auch $F(x) = \int f(x)\,dx$ und spricht vom *unbestimmten Integral* von f.

Substitution in Lebesgue-Integralen:

$$\int_\alpha^\beta g(x(t))\dot{x}(t)\,dt = \int_{x(\alpha)}^{x(\beta)} g(x)\,dx \text{ oder}$$

– falls $x = x(t)$ eindeutig nach t auflösbar ist:

$$t = t(x) - \int_a^b g(x)\,dx = \int_{t(a)}^{t(b)} g[x(t)]\dot{x}(t)\,dt.$$

13.4.2.4. Simpsonsche Regel. Zur näherungsweisen Berechnung von $\int\limits_a^b f(x)\,dx$ wird das Intervall $a \leqq x \leqq b$ durch $x_k = a + hk$ $(k = 0, 1, \ldots, 2n$; $h =$ *Schrittweite*$)$ in eine gerade Anzahl $2n$ von Teilintervallen zerlegt, deren Längen jeweils gleich $h = (b - a)/2n$ sind. In einem ersten Schritt werden für das Integral $\int\limits_{x_{2\nu}}^{x_{2\nu+2}} f(x)\,dx$ dadurch drei vorläufige grobe Näherungen hergestellt, indem $f(x)$ durch konstante Funktionen ersetzt wird, deren Werte gleich den Funktionswerten von $f(x)$ an den Stellen $x = x_{2\nu}$ bzw. $x = x_{2\nu+1}$ bzw. $x = x_{2\nu+2}$ sind. Mit den erhaltenen Näherungen $2hf(x_{2\nu})$, $2hf(x_{2\nu+1})$, $2hf(x_{2\nu+2})$ soll durch die Linearkombination $F_\nu = 2h[c_1f(x_{2\nu}) + c_2f(x_{2\nu+1}) + c_3f(x_{2\nu+2})]$ eine möglichst gute Näherung für dieses Integral durch geeignete Wahl der Konstanten c_1, c_2, c_3 hergestellt werden. Hierzu wird mittels der Taylor-Entwicklung von $f(x)$ an der Stelle $x = x_{2\nu+1}$ einerseits durch gliedweise Integration eine Darstellung des Integrals durch eine Potenzreihe nach Potenzen von h hergestellt und andererseits eine ebensolche Potenzreihe für den obigen Näherungswert F_ν gewonnen. Die Konstanten werden danach derart festgelegt, daß beide erhaltenen Ent-

wicklungen bis zu möglichst hohen Potenzen von h übereinstimmen (*Taylor-Abgleich*). Es ergibt sich $F_\nu = [f(x_{2\nu}) + 4f(x_{2\nu+1}) + f(x_{2\nu+2})]h/3$ und eine Übereinstimmung bis zum Glied mit h^4. Hieraus folgt $\int_a^b f(x)\mathrm{d}x = F + R$ mit dem Näherungswert $F = [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + 2f(x_4) + \ldots + 2f(x_{2n-2}) + 4f(x_{2n-1}) + f(x_{2n})]h/3$ und dem Restglied $R = -(1/90)[(b-a)/2n]^5 f^{(4)}(\zeta)$, wobei aber von ζ nur $a < \zeta < b$ bekannt ist. Will man die Ableitungsbildung bei der Ermittlung von R vermeiden, so kann – falls $2n$ durch 4 teilbar ist – R *näherungsweise* durch $(1/15)(F-G)$ angegeben werden, wobei sich G aus der Simpson-Regel mit der *doppelten* Schrittweite $2h$ ergibt: $G = (2h/3)[f(x_0) + 4f(x_2) + 2f(x_4) + 4f(x_6) + 2f(x_8) + \ldots + 2f(x_{2n-4}) + 4f(x_{2n-2}) + f(x_{2n})]$ (F = „feine" Näherung, G = „grobe" Näherung).

13.4.3. Integration von reellen Funktionen von zwei reellen unabhängigen Variablen

13.4.3.1. Lebesgue-Stieltjes-Maße. Es sei F eine reelle Funktion, die als Definitionsbereich die Menge $\Omega = R^2$ aller reellen Zahlenpaare besitzt und sowohl bezüglich x als auch bezüglich y einerseits monoton nicht-fallend und andererseits linksseitig stetig ist. Es gibt ein möglichst umfangreiches System B_F, das aus Teilmengen von R^2 – den *bezüglich F Lebesgue-Stieltjes-meßbaren Mengen* – besteht und Definitionsbereich des *bezüglich F Lebesgue-Stieltjes-Maßes* μ_F ist, falls man von F darüber hinaus voraussetzt, daß die rechte Seite der Gleichung $\mu_F(\{(x,y) \mid x_1 \leqq x \leqq x_2, y_1 \leqq y \leqq y_2\}) = F(x_2,y_2) - F(x_2,y_1) - F(x_1,y_2) + F(x_1,y_1)$ niemals negativ ausfällt. Ist speziell $F(x,y) = F_0(x,y) = xy$, so ergeben sich definitionsgemäß die *Lebesgue-meßbaren Mengen* mit dem dazugehörigen *Lebesgue-Maß*, wobei $\mu_{F_0}(\{(x,y) \mid x_1 \leqq x \leqq x_2, y_1 \leqq y \leqq y_2\})$ gleich dem gewöhnlichen Rechteckinhalt $(x_2 - x_1)(y_2 - y_1)$ ist.

Bsp. 13.4.3–1. Das Paar (X, Y) von Zufallsgröße X und Y über dem gemeinsamen Wahrscheinlichkeitsraum $[\Omega, B, P]$ (vgl. 13.3.1.7.) heißt *(reeller zweidimensionaler) zufälliger Vektor*. Der Funktionswert $F(x,y)$ der zugehörigen *Verteilungsfunktion* F ist gleich der Wahrscheinlichkeit derjenigen Teilmenge von Ω, für deren Elemente ω sowohl $X(\omega) < x$ als auch $Y(\omega) < y$ gilt. Dieses $F = F(x,y)$ definiert ein Lebesgue-Stieltjes-Maß und hat darüber hinaus die Eigenschaften $\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$ und $\lim F(x,y) = 1$, wobei im letzten Limes sowohl x als auch y nach ∞ streben. Ein Beispiel ist die *zweidimensionale Normalverteilung* für die $F(x,y) = \int_{-\infty}^{x}\left(\int_{-\infty}^{y} f(x,y)\mathrm{d}y\right)\mathrm{d}x$ gilt mit der *Wahrscheinlichkeitsdichte* $f(x,y) = \{2\pi\sigma_x\sigma_y\}^{-1}(1-\varrho^2)^{-1/2} \exp\{-[2(1-\varrho^2)]^{-1}[((x-\mu_x)\sigma_x)^2 - 2\varrho(x-\mu_x)(y-\mu_y)/(\sigma_x\sigma_y) + ((y-\mu_y)/\sigma_y)^2]\}$. Während die Normalverteilung $N(\mu, \sigma^2)$ (vgl. Bsp. 13.3.1–4) nur zwei Parameter μ und σ^2 aufweist, liegen hier die fünf Parameter μ_x, μ_y, σ_x, σ_y, ϱ vor.

13.4.3.2. Lebesgue-Stieltjes-Integral. Die Definition des *Lebesgue-Stieltjes-Integrals* ist mittels 13.4.3.1. dem Abschnitt 13.4.1.3. zu entnehmen. Liegt der Integrationsbereich $\Omega' = B$ vor, so schreibt man für $\int_{\Omega'} g(x,y)\mathrm{d}\mu_F(x,y)$ gern $\int_B g(x,y)\mathrm{d}F(x,y)$ oder auch $\iint_B g(x,y)\mathrm{d}F(x,y)$. Im Spezialfall $F(x,y) = F_0(x,y) = xy$ ergeben sich die *Lebesgue-Integrale* und man schreibt für $\int_B g(x,y)\mathrm{d}F_0(x,y)$ kurz $\int_B g(x,y)\mathrm{d}b$ oder $\int_B g(x,y)\mathrm{d}x\,\mathrm{d}y$ oder ersetzt auch in diesen Formeln $\int_B$ durch $\iint_B$.

Bsp. 13.4.3–2. Unter der *Kovarianz* des zufälligen Vektors (X, Y) (vgl. Bsp. 13.4.3–1) versteht man $\mathrm{cov}(X,Y) = \iint_{R^2} (x - EX)(y - EY)\mathrm{d}F(x,y)$.

13.4.3.3. Produktmaß. Ist μ_{F_1} mit $F_1 = F_1(x)$ ein bezüglich F_1 Lebesgue-Stieltjes-Maß auf der reellen x-Achse und μ_{F_2} mit $F_2 = F_2(y)$ ein solches auf der y-Achse, so gibt es ein möglichst umfangreiches System von Teilmengen des R^2, das Definitionsbereich eines *Produktmaßes* μ_F mit $F = F(x,y) = F_1(x)F_2(y)$ ist, wobei $\mu_F(\{(x,y) \mid x_1 \leqq x \leqq x_2, y_1 \leqq y \leqq y_2\}) = \mu_{F_1}(\{x \mid x_1 \leqq x \leqq x_2\})\mu_{F_2}(\{y \mid y_1 \leqq y \leqq y_2\})$ gilt. Das durch $F(x,y) = F_0(x,y) = xy$ definierte Lebesgue-Maß ist ein Produktmaß. Das durch die Verteilungsfunktion $F(x,y)$ aus Bsp. 13.4.3–1 definierte Maß ist genau dann ein Produktmaß, falls X und Y *stochastisch unabhängig* sind. Es gilt dann $F(x,y) = F_1(x)F_2(y)$ mit $F_1(x) = \lim_{y\to\infty} F(x,y)$ und $F_2(y) = \lim_{x\to\infty} F(x,y)$.

13.4.3.4. Zurückführen auf ein iteriertes Integral. *Satz von Fubini*. Ist $g = g(x,y)$, $(x,y) \in R^2$ bezüglich des Produktmaßes μ_F $[F(x,y) = F_1(x)F_2(y)]$ integrierbar, so gilt $\iint_{R^2} g(x,y)\mathrm{d}[F_1(x)F_2(y)] = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} g(x,y)\mathrm{d}F_1(x)\right\}\mathrm{d}F_2(y) = \int_{-\infty}^{\infty}\left\{\int_{-\infty}^{\infty} g(x,y)\mathrm{d}F_2(y)\right\}\mathrm{d}F_1(x)$. Die beiden letzten Integrale sind *iterierte Integrale*; man berechnet zunächst die inneren und danach die äußeren Integrale.

Bsp. 13.4.3–3. Hat der Integrationsbereich von $g(x,y)$ die Struktur $\Omega' = B = \{(x,y) \mid a \leqq x \leqq b, y_1(x) \leqq y \leqq y_2(x)\}$ bzw. $\Omega' = B = \{(x,y) \mid x_1(y) \leqq x \leqq x_2(y), c \leqq y \leqq d\}$, so gilt für das Lebesgue-Integral $\iint_B g(x,y)\mathrm{d}b = \int_a^b\left\{\int_{y_1(x)}^{y_2(x)} g(x,y)\mathrm{d}y\right\}\mathrm{d}x$ bzw. $\iint_B g(x,y)\mathrm{d}b = \int_c^d\left\{\int_{x_1(y)}^{x_2(y)} g(x,y)\mathrm{d}x\right\}\mathrm{d}y$.

13.4.3.5. Variablentransformation in Flächenintegralen. Durch $x = x(u,v), y = y(u,v)$ werde der Integrationsbereich von $g(x,y)$ eineindeutig auf $\tilde{B}$ der (u,v)-Ebene abgebildet, wobei gegebenenfalls durch Vertauschen von u und v gesichert werde, daß die Funktionaldeterminante $\frac{\partial(x,y)}{\partial(u,v)} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial y}{\partial u}\frac{\partial x}{\partial v}$ größer als null ist. Dann gilt $\iint_B g(x,y)\mathrm{d}b = \iint_{\tilde{B}} g(x(u,v),y(u,v))\frac{\partial(x,y)}{\partial(u,v)}\mathrm{d}\tilde{b}$.

Liegen speziell ebene Polarkoordinaten $(u,v) = (r_\perp, \varphi)$ (d. h. Zylinderkoordinaten $(r_\perp, \varphi, z)$ im Fall $z = 0$, vgl. Bsp. 13.2.5–1) vor, so ist die Funktionaldeterminante gleich $r_\perp$.

13.4.4. Integration von reellen Funktionen von drei reellen unabhängigen Variablen

Die Übertragung von 13.4.3. wird hier nicht ausgeführt. Lediglich an Bsp. 13.4.3–3 angeknüpft.

13.4.4.1. Zerlegung des Integrationsbereiches in Säulen. Hat der Integrationsbereich von $g(x,y,z)$ die Struktur $\Omega' = V = \{(x,y,z) \mid (x,y) \in B, z_1(x,y) \leqq z \leqq z_2(x,y)\}$, so gilt für das Lebesgue-Integral

$$\iiint\limits_V g(x,y,z)\,\mathrm{d}V = \iint\limits_B \left\{ \int\limits_{z_1(x,y)}^{z_2(x,y)} g(x,y,z)\,\mathrm{d}z \right\} \mathrm{d}b.$$

13.4.4.2. Zerlegung des Integrationsbereiches in Schichten. Hat der Integrationsbereich von $g(x,y,z)$ die Struktur $\Omega' = V = \{(x,y,z) \mid (y,z) \in B = B(x), a \leqq x \leqq b\}$, so gilt für das Lebesgue-Integral

$$\iiint\limits_V g(x,y,z)\,\mathrm{d}V = \int\limits_a^b \left\{ \iint\limits_{B(x)} g(x,y,z)\,\mathrm{d}b(x) \right\} \mathrm{d}x.$$

Entsteht speziell V durch Rotation der Kurve $y = y(x)$ $(a \leqq x \leqq b)$ um die x-Achse und hängt darüber hinaus $g(x,y,z)$ nur von x ab, also $g(x,y,z) = g(x)$, so ist

$$\iiint\limits_V g(x)\,\mathrm{d}V = \pi \int\limits_a^b g(x)\,(y(x))^2\,\mathrm{d}x.$$

13.4.4.3. Variablentransformation in Raumintegralen. Die Übertragung von 13.4.3.5. führt zu

$$\iiint\limits_V g(x,y,z)\,\mathrm{d}V$$

$$= \iiint\limits_V g[x(u,v,w), y(u,v,w), z(u,v,w)] \frac{\partial(x,y,z)}{\partial(u,v,w)}\,\mathrm{d}V \quad \text{mit}$$

$$\frac{\partial(x,y,z)}{\partial(u,v,w)} = \frac{\partial x}{\partial u}\frac{\partial(y,z)}{\partial(v,w)} - \frac{\partial y}{\partial u}\frac{\partial(x,z)}{\partial(v,w)} + \frac{\partial z}{\partial u}\frac{\partial(x,y)}{\partial(v,w)} > 0.$$

Im Fall von Zylinderkoordinaten $(r_\perp, \varphi, z)$ bzw. Kugelkoordinaten (r, ϑ, φ) (vgl. Bsp. 13.2.5–1, 13.2.5–2) ist die Funktionaldeterminante gleich $r_\perp$ bzw. $r^2 \sin\vartheta$.

13.4.5. Kurvenintegrale

Der Integrationsbereich ist eine Kurve $\boldsymbol{r}(t) = x_k(t)\boldsymbol{e}_k$, $\alpha \leqq t \leqq \beta$ (vgl. 13.3.1.3.).

13.4.5.1. Kurvenintegrale erster Art. *Das Kurvenintegral erster Art* $\int\limits_C g\,\mathrm{d}s$ kann mit dem Bogenlängendifferential $\mathrm{d}s = |\mathrm{d}\boldsymbol{r}| = \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}\,\mathrm{d}t$ (vgl. 13.3.3.4.) durch $\int\limits_C g\,\mathrm{d}s = \int\limits_\alpha^\beta g(x_1(t), x_2(t), x_3(t)) \sqrt{\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2}\,\mathrm{d}t$ definiert und berechnet werden.

Bsp. 13.4.5–1. Die Formel $\boldsymbol{r}_S = \frac{1}{l}\int\limits_C \boldsymbol{r}\,\mathrm{d}s$ für den Ortsvektor $\boldsymbol{r}_S$ des *(geometrischen) Schwerpunktes* der Kurve C ($l = \int\limits_C \mathrm{d}s$ ist die *Bogenlänge* von C) zeigt, daß der Integrand auch ein Vektor sein kann.

13.4.5.2. Kurvenintegrale zweiter Art. Im Vektorfeld $\boldsymbol{F}(\boldsymbol{r}) = \boldsymbol{F}(x_1, x_2, x_3) = F_k(x_1, x_2, x_3)\boldsymbol{e}_k$ (vgl. 13.3.3.7.) kann das *Kurvenintegral zweiter Art* $\int\limits_C \boldsymbol{F}(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{r}$ mit dem vektoriellen Kurvendifferential $\mathrm{d}\boldsymbol{r} = \dot{x}_k(t)\boldsymbol{e}_k\mathrm{d}t$ (vgl. 13.3.3.4.) durch

$$\int\limits_\alpha^\beta F_k[x_1(t), x_2(t), x_3(t)]\dot{x}_k(t)\,\mathrm{d}t$$

(Einsteinsche Summationsverabredung!) definiert und berechnet werden. Die beiden Kurvenintegrale $\int\limits_{C_1} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$ und $\int\limits_{C_2} \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{r}$, wobei die Anfangs- und Endpunkte von C_1 und C_2 übereinstimmen, sind einander gleich, wenn es bei festgehaltenem Anfangs- und Endpunkt einen stetigen Deformationsprozeß gibt, der C_1 in C_2 überführt, wobei in den überstrichenen Punkten stets $\mathbf{rot}\,\boldsymbol{F} = 0$ gilt. In diesem Fall ist $\boldsymbol{F}$ ein *Gradientenfeld* $\boldsymbol{F} = \mathbf{grad}\,U(x_1, x_2, x_3)$ und der gemeinsame Integralwert ist gleich der Differenz der Funktionswerte von U am End- und am Anfangspunkt der Kurven. Ist $\boldsymbol{F}$ speziell ein Kraftfeld, so wird in der Mechanik $-U$ als *Potential* bezeichnet.

13.4.6. Oberflächenintegrale

Der Integrationsbereich ist eine Fläche $\boldsymbol{r}(u,v) = x_k(u,v)\boldsymbol{e}_k$, $(u,v) \in B$ (vgl. 13.3.1.3.).

13.4.6.1. Oberflächenintegrale erster Art. Mit dem *vektoriellen Oberflächendifferential* $\mathrm{d}\boldsymbol{A} = \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\,\mathrm{d}u\,\mathrm{d}v$ (vgl. 13.3.3.5.), das senkrecht auf der jeweiligen Tangentialebene der Fläche A steht, ergibt sich das *(skalare) Oberflächendifferential* $\mathrm{d}A = |\mathrm{d}\boldsymbol{A}| = \sqrt{EG - F^2}\,\mathrm{d}u\,\mathrm{d}v$ mit $E = \boldsymbol{r}_{,u} \cdot \boldsymbol{r}_{,u}$, $F = \boldsymbol{r}_{,u} \cdot \boldsymbol{r}_{,v}$ und $G = \boldsymbol{r}_{,v} \cdot \boldsymbol{r}_{,v}$. Somit kann das *Oberflächenintegral erster Art* entsprechend

$$\iint\limits_A g\,\mathrm{d}A = \iint\limits_B g[x_1(u,v), x_2(u,v), x_3(u,v)]\sqrt{EG - F^2}\,\mathrm{d}u\,\mathrm{d}v$$

definiert und berechnet werden.

13.4.6.2. Oberflächenintegrale zweiter Art. Das *Oberflächenintegral zweiter Art* $\iint\limits_A \boldsymbol{F}(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{A}$ kann mit dem vektoriellen Oberflächendifferential $\mathrm{d}\boldsymbol{A}$ (vgl. 13.4.6.1.) durch $\iint\limits_A \boldsymbol{F}(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{A} = \iint\limits_B [\boldsymbol{F}(\boldsymbol{r})\boldsymbol{r}_{,u}\boldsymbol{r}_{,v}]\,\mathrm{d}u\,\mathrm{d}v$ (vgl. 13.2.5.1., Spatprodukt) definiert und berechnet werden.

13.4.6.3. Der Integralsatz von Gauss. Bezeichnet ∂V die Oberfläche eines räumlichen Bereiches V und ist das vektorielle Oberflächendifferential $\mathrm{d}\boldsymbol{A}$ stets in das Äußere von V gerichtet, so lautet der *Integralsatz von Gauss*

$$\iiint\limits_V \operatorname{div} \boldsymbol{F}(\boldsymbol{r})\,\mathrm{d}V = \oiint\limits_{\partial V} \boldsymbol{F}(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{A}$$

und damit in kartesischen Koordinaten $\iiint\limits_V F_{k,k}\,\mathrm{d}V = \oiint\limits_{\partial V} F_k\,\mathrm{d}A_k$ $(\mathrm{d}\boldsymbol{A} = \mathrm{d}A_k\boldsymbol{e}_k)$. Der Kreis im Integralzeichen weist auf die *geschlossene* Oberfläche ∂V hin. Eine Verallgemeinerung entsteht dadurch, indem man in der letzten Formel F_k durch die Koordinaten eines Tensorfeldes n-ter Stufe $(n = 0, 2, 3, 4, \ldots)$ ersetzt. Es werde dies in der folgenden Formel durch $(\ldots)$ angedeutet:

$$\iiint\limits_V (\ldots)_{,k}\,\mathrm{d}V = \oiint\limits_{\partial V} (\ldots)\,\mathrm{d}A_k.$$

Bsp. 13.4.6–1. In Fortführung von Bsp. 13.2.6–2 ergibt sich mit 13.2.6.2. für die Gesamtkraft, die auf die Oberfläche ∂V eines räumlichen Bereiches V in einem belasteten Kontinuum wirkt:

$$\oiint\limits_{\partial V} \sigma \boldsymbol{n}\,\mathrm{d}A = \oiint\limits_{\partial V} \sigma\,\mathrm{d}\boldsymbol{A} = \oiint\limits_{\partial V} \sigma_{kl}\,\mathrm{d}A_l\boldsymbol{e}_k = \iiint\limits_V \sigma_{kl,l}\,\mathrm{d}V\boldsymbol{e}_k.$$

13.4.6.4. Der Integralsatz von Stokes. Bezeichnet C die Randkurve eines Flächenstückes A und sind einerseits C und andererseits $\mathrm{d}\boldsymbol{A}$ derart orientiert, daß ein Wandern – im Sinne der Orientierung – längs C verbunden mit einem Fortschreiten in Richtung von $\mathrm{d}\boldsymbol{A}$ eine Rechtsschraubung ergibt, so lautet der *Integralsatz von Stokes*

$\iint_A \operatorname{rot} F(r)\cdot \mathrm{d}A = \oint_C F(r)\cdot \mathrm{d}r$ und damit in kartesischen Koordinaten $\iint_A \varepsilon_{klm}F_{l,k}\mathrm{d}A_m = \oint_C F_l \mathrm{d}x_l$ $(\mathrm{d}r = \mathrm{d}x_l e_l)$. Analog zu 13.4.6.3. kann eine Verallgemeinerung durch $\iint_A \varepsilon_{klm}(\ldots)_{,k}\mathrm{d}A_m = \oint_C (\ldots)\mathrm{d}x_l$ angegeben werden.

Bsp. 13.4.6–2. Mit dem Tensorfeld nullter Stufe $U = U(x_1, x_2, x_3)$ ist $\oint_C U\mathrm{d}r = \oint_C U\mathrm{d}x_l e_l = \iint_A \varepsilon_{klm} U_{,k}\mathrm{d}A_m e_l = -\iint_A \operatorname{grad} U \times \mathrm{d}A$.

Im Integralsatz von Stokes ist $\varepsilon_{klm}\mathrm{d}A_m$ wegen $\mathrm{d}A_m = \varepsilon_{pqm}\frac{\partial x_p}{\partial u}\frac{\partial x_q}{\partial v}\mathrm{d}u\mathrm{d}v$ und $\varepsilon_{klm}\varepsilon_{pqm} = \delta_{kp}\delta_{lq} - \delta_{kq}\delta_{lp}$ (vgl. 13.2.5.2., Zerlegungssatz) gleich den Koordinaten

$$\mathrm{d}A_{kl} = \left(\frac{\partial x_k}{\partial u}\frac{\partial x_l}{\partial v} - \frac{\partial x_l}{\partial u}\frac{\partial x_k}{\partial v}\right)\mathrm{d}u\mathrm{d}v$$

des *antimetrischen Oberflächendifferentialtensors zweiter Stufe*.

13.4.7. Integration von Funktionen, deren Funktionswerte komplex sind

Sind in 13.4.1.3. die Funktionswerte von g komplex, gilt also $g = \operatorname{Re} g + \mathrm{i}\operatorname{Im} g$ (vgl. 13.1.3.3.), so definiert man $\int_\Omega g\mathrm{d}\mu = \int_\Omega \operatorname{Re} g\mathrm{d}\mu + \mathrm{i}\int_\Omega \operatorname{Im} g\mathrm{d}\mu$.

13.4.7.1. Integration komplexer Funktionen einer komplexen unabhängigen Variablen. Ist $z(t) = x(t) + \mathrm{i}y(t)$ $(\alpha \leqq t \leqq \beta)$ eine Parameterdarstellung der in der Gaußschen Zahlenebene liegenden Kurve C und $w(z) = u(x,y) + \mathrm{i}v(x,y)$ mit $z = x + \mathrm{i}y$ (vgl. 13.3.1.4.), so definiert man

$$\int_C w(z)\mathrm{d}z = \int_\alpha^\beta w(z(t))\dot{z}(t)\mathrm{d}t = \int_\alpha^\beta (u + \mathrm{i}v)(\dot{x} + \mathrm{i}\dot{y})\mathrm{d}t$$
$$= \int_\alpha^\beta (u\dot{x} - v\dot{y})\mathrm{d}t + \mathrm{i}\int_\alpha^\beta (u\dot{y} + v\dot{x})\mathrm{d}t,$$

wobei man in $u = u(x,y)$ und $v = v(x,y)$ für x und y jeweils $x = x(t), y = y(t)$ einzusetzen hat.

13.4.7.2. Residuensatz. Riemannsche Zahlenkugel (RZK). Man denkt sie sich im Punkt $z = 0$ auf die Gaußsche Zahlenebene (GZE) gesetzt. Verbindet man den Nordpol der RZK jeweils mit den Punkten der GZE durch eine Gerade, so durchstößt diese die RZK jeweils in einem Punkt. Es entsteht auf diese Weise eine eineindeutige – *stereografische Projektion* genannte – Abbildung der GZE auf die im Nordpol punktierte RZK. Durch Hinzunahme des (einzigen) Punktes ∞ geht die GZE in die *Vollebene* (*funktionentheoretische Ebene*, *geschlossene Zahlenebene*) über. Man definiert, daß bei der stereografischen Projektion $z = \infty$ dem Nordpol der RZK zugeordnet wird.

Residuum. Ist die Funktion $w = w(z)$ in einer punktierten Umgebung von $z = z_0 \neq \infty$, also für alle z mit $0 < |z - z_0| < r$ holomorph (vgl. 13.3.3.9.), so ist sie dort in eine *Laurent-Reihe* (vgl. 13.3.4.5.) $w(z) = \sum_{\nu=-\infty}^{\infty} c_\nu (z - z_0)^\nu = \ldots + c_{-2}(z - z_0)^{-2} + c_{-1}(z - z_0)^{-1} + c_0 + c_1(z - z_0) + \ldots$ entwickelbar. Der Koeffizient c_{-1} ist das *Residuum* von $w = w(z)$ *an der Stelle* $z = z_0$: $c_{-1} = \operatorname{Res}_{z=z_0} w(z)$. Ist $w = w(z)$ in einer Umgebung von $z = z_0 = \infty$, d. h. für $|z| > R$ holomorph, so ist eine Laurent-Reihenentwicklung an der Stelle $z_0 = \infty$ möglich (vgl. 13.3.4.5.); mit $t = z^{-1}$ ist

$$w(z) = g(t) = \sum_{\nu=-\infty}^{\infty} c_\nu t^\nu = \sum_{\nu=-\infty}^{\infty} c_\nu (1/z)^\nu$$
$$= \ldots + c_{-2}z^2 + c_{-1}z + c_0 + c_1(1/z) + c_2(1/z)^2 + \ldots.$$

Mit dieser Bezeichnung gilt für das Residuum von $w(z)$ an der Stelle $z = \infty$ die Formel $\operatorname{Res}_{z=\infty} w(z) = -c_1$.

Die in der GZE liegende geschlossene Kurve C, auf der nicht der Punkt ∞ liege – also ist beispielsweise die Achse des Reellen nicht als Kurve C brauchbar – zerlege die Vollebene in zwei Gebiete G_1 und G_2, wobei der Punkt ∞ nicht zu G_1, jedoch zu G_2 gehöre.

Residuensatz, *falls* ∞ *nicht beteiligt ist*. Ist $w = w(z)$ in G_1 vereinigt mit C stetig und in G_1 holomorph, ausgenommen an den Stellen $z = z_\nu \in G_1$ $(\nu = 1, \ldots, n)$, so lautet der Residuensatz

$$\oint_C w(z)\mathrm{d}z = 2\pi\mathrm{i}\sum_{\nu=1}^{n} \operatorname{Res}_{z=z_\nu} w(z),$$

wobei die Kurve C derart zu orientieren ist, daß bei ihrer Durchwanderung gemäß ihrer Orientierung G_1 stets zur Linken liegt.

Residuensatz, *falls* ∞ *beteiligt ist*. Ist $w = w(z)$ in G_2 vereinigt mit C stetig und in G_2 holomorph, ausgenommen an den Stellen $z = z_\nu \in G_2$ $(z_\nu \neq \infty, \nu = 1, \ldots, n)$ und ∞, so lautet der Residuensatz

$$\oint_C w(z)\mathrm{d}z = 2\pi\mathrm{i}\left\{\sum_{\nu=1}^{n} \operatorname{Res}_{z=z_\nu} w(z) + \operatorname{Res}_{z=\infty} w(z)\right\},$$

wobei die Kurve C derart zu orientieren ist, daß bei ihrer Durchwanderung gemäß ihrer Orientierung G_2 stets zur Linken liegt.

Erfüllt $w(z)$ die Voraussetzungen für jeden der oben genannten Residuensätze, so unterscheiden sich – infolge der Orientierung – die von den Sätzen gelieferten Integrale um den Faktor -1.

Bsp. 13.4.7–1. Ist C der im mathematisch positiven Sinn orientierte Kreis $|z - z_0| = r$, so ist $\oint_C \frac{\mathrm{d}z}{z - z_0} = 2\pi\mathrm{i}$. Der Wert des Integrals ändert sich nicht, wenn man den Kreis stetig deformiert und dabei den Punkt $z = z_0$ nicht überstreicht.

Bsp. 13.4.7.–2. Ist $w = w(z)$ in G_1 (siehe oben) sogar überall holomorph, so ergibt sich mit dem Residuensatz die *Cauchysche Integralformel*

$$w(z) = \frac{1}{2\pi\mathrm{i}}\oint_C \frac{w(\zeta)}{\zeta - z}\mathrm{d}\zeta \quad (z \in G_1).$$

Lemma von Jordan. C_0 sei der in der oberen Halbebene $(\operatorname{Im} z \geqq 0)$ liegende Halbkreis mit dem Radius R und dem Mittelpunkt 0. Strebt $w(z)$ ($w(z)$ stetig) in der oberen Halbebene für $z \to \infty$ gleichmäßig gegen 0, d. h. ist $\lim_{R\to\infty} \max_{0\leqq\varphi\leqq\pi} |w(R\mathrm{e}^{\mathrm{i}\varphi})| = 0$, so gilt $\lim_{R\to\infty} \int_{C_0} w(z)\mathrm{e}^{\mathrm{i}tz}\mathrm{d}z = 0$, falls $t > 0$ ist. Liegt der Halbkreis C_u mit dem Radius R und dem Mittelpunkt 0 in der unteren Halbebene, so gilt unter der entsprechenden Voraussetzung die obige Behauptung mit $t < 0$. Ein analoger Sachverhalt liegt für Integrale der Gestalt $\int_{C_l} w(z)\mathrm{e}^{tz}\mathrm{d}z$ vor, falls C_l in der linken Halbebene $(\operatorname{Re} z \leqq 0)$ liegt und $t > 0$ ist, und für $\int_{C_r} w(z)\mathrm{e}^{tz}\mathrm{d}z$, falls C_r in der rechten Halbebene $(\operatorname{Re} z \geqq 0)$ liegt und $t < 0$ ist. Die Behauptung ist jeweils auch im Fall $t = 0$ gesichert, falls man in der Voraussetzung $\lim_{R\to\infty} \max |w(R\mathrm{e}^{\mathrm{i}\varphi})| = 0$ zu $\lim_{R\to\infty} \max |Rw(R\mathrm{e}^{\mathrm{i}\varphi})| = 0$ verschärft.

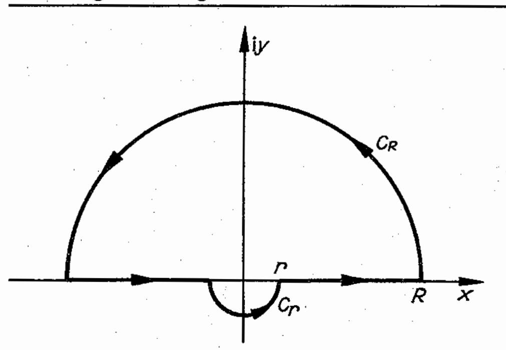


Abb. 13.4.7–1 Integrationsweg C bei Bsp. 13.4.7–3

Bsp. 13.4.7–3. $J = \int\limits_0^\infty \frac{\sin(tx)}{x} dx \, (t > 0)$. Wegen der Geradheit des Integranden $g(x) = \frac{\sin(tx)}{x} [g(-x) = g(x)]$ ist $2J = \int\limits_{-\infty}^{\infty} \frac{\sin(tx)}{x} dx$. Als Vorbereitung zur Anwendung des Residuensatzes in Verbindung mit dem Jordanschen Lemma schreibt man $2J = \operatorname{Im}\left\{ \lim\limits_{R\to\infty} \lim\limits_{r\to+0} \left(\int\limits_{-R}^{-r} \frac{e^{itx}}{x} dx + \int\limits_{r}^{R} \frac{e^{itx}}{x} dx \right) \right.$ $= \operatorname{Im} \lim\limits_{R\to\infty} \lim\limits_{r\to+0} \oint\limits_C \frac{e^{itz}}{z} dz - \lim\limits_{r\to+0} \int\limits_{C_r} \frac{e^{itz}}{z} dz - \left. \lim\limits_{R\to\infty} \int\limits_{C_R} \frac{e^{itz}}{z} dz \right\}$ mit den orientierten Halbkreisen $C_r = \{z = re^{i\varphi} \mid \pi \leqq \varphi \leqq 2\pi\}$ und $C_R = \{z = Re^{i\varphi} \mid 0 \leqq \varphi \leqq \pi\}$ und der geschlossenen Kurve $C = \{x \mid -R \leqq x \leqq -r\} \cup C_r \cup \{x \mid r \leqq x \leqq R\} \cup C_R$ (Abb. 13.4.7–1). Die Einführung von C_r ist erforderlich, weil sonst die Polstelle $z = 0$ von $z^{-1}e^{itz}$ auf dem Integrationsweg liegen würde. Der Residuensatz ergibt $\oint\limits_C \frac{e^{itz}}{z} dz = 2\pi i \operatorname{Res}\limits_{z=0} \frac{e^{itz}}{z} = 2\pi i$. Das Lemma von Jordan liefert $\lim\limits_{R\to\infty} \int\limits_{C_R} \frac{e^{itz}}{z} dz = 0$ und mit 13.4.7.2. erhält man $\lim\limits_{r\to+0} \int\limits_{C_r} \frac{e^{itz}}{z} dz = \lim\limits_{r\to+0} i \int\limits_{\pi}^{2\pi} e^{itr(\cos\varphi + i\sin\varphi)} d\varphi = i\pi$. Also ist $J = \pi/2$.

Bsp. 13.4.7–5. Beim Arbeiten mit der Laplace-Transformation (vgl. Bsp. 13.1.5–7) wird man die Rücktransformation (vgl. 13.3.1.) möglichst mittels Tafelwerken vornehmen. Ist dies nicht möglich, so steht die *Umkehrformel* $f(t) = \frac{1}{2\pi i} \lim\limits_{R\to\infty} \int\limits_{c-iR}^{c+iR} e^{tp} F(p) dp$ [c ist reell, der Integrationsweg verläuft parallel zur Achse des Imaginären und liegt im Holomorphiebereich $\operatorname{Re} p > a$ von $F(p)$] zur Verfügung. Behandelt wird das Beispiel $F(p) = p(p^2 + \omega^2)^{-1} \exp(-\sqrt{p}\,\alpha)$, $\omega > 0, \alpha > 0$; von der Wurzel ist der Hauptwert gemeint. Da $F(p)$ für $\operatorname{Re} p > 0$ holomorph ist, kann $c > 0$ gewählt werden. Um den Residuensatz in Verbindung mit dem Lemma von Jordan anwenden zu können, wird man zunächst die geschlossene Kurve $C = \{z = c + i\tau \mid -R \leqq \tau \leqq R\} \cup C_l$ mit dem Halbkreis $C_l = \{z = c + Re^{i\varphi} \mid (\pi/2) \leqq \varphi \leqq 3\pi/2\}$ betrachten. An den Nullstellen des Nenners $p^2 + \omega^2$, also bei $p = \pm i\omega$ liefern die Laurent-Reihenentwicklungen von e^{tp} $(p^2 + \omega^2)^{-1} \exp(-\sqrt{p}\,\alpha)$ (vgl. 13.3.4.5.) die Residuen $(1/2) \exp(ti\omega - \alpha\sqrt{\omega}\sqrt{i})$ und $(1/2) \exp(-ti\omega - \alpha\sqrt{\omega}\sqrt{-i})$.

Die Summe dieser Residuen ist wegen $\sqrt{i} = (1 + i)/\sqrt{2}$ und $\sqrt{-i} = (1 - i)/\sqrt{2}$ gleich $\exp(-\alpha\sqrt{\omega/2}) \cos(\omega t - \alpha\sqrt{\omega/2})$. Da hier die Voraussetzungen des Lemmas von Jordan erfüllt sind, ist es bei oberflächlicher Betrachtung naheliegend – gestützt auf den Residuensatz – das obige Ergebnis mit dem gesuchten $f(t)$ mit $t > 0$ zu identifizieren. Leider sind jedoch die Voraussetzungen des Residuensatzes nicht erfüllt, da der Hauptwert von $\sqrt{p}$ längs der negativ-reellen Achse springt (vgl. 13.3.4.6., Hauptwert der Potenz; 13.4.7.4., Riemannsche Fläche des komplexen Logarithmus) und also dort keine Holomorphie vorliegen kann. Um den Residuensatz anwenden zu können, wird der obige geschlossene Integrationsweg C durch folgende Kurven $C_g = C_o \cup C_r \cup C_u$ ergänzt (Abb. 13.4.7–2): C_o erstreckt sich längs des oberen Ufers der negativ-reellen Achse vom Schnittpunkt des Halbkreises C_l mit der negativ-reellen Achse bis zum Punkt $-r (0 < r < c)$. C_r ist der im mathematisch negativen Sinn orientierte Kreis mit dem Mittelpunkt 0 und dem Radius r, der sich von $-r$ auf dem oberen Ufer bis zu $-r$ auf dem unteren Ufer erstreckt. C_u verläuft längs des unteren Ufers der negativ-reellen Achse von $-r$ bis zum Schnittpunkt des Halbkreises C_l mit der negativ-reellen Achse. Auf den Integrationsweg $C \cup C_g$ ist nun der Residuensatz anwendbar, so daß mit dem obigen Ergebnis $f(t) = \exp(-\alpha\sqrt{\omega/2}) \cos(\omega t - \sqrt{\omega/2})$ $- \lim\limits_{R\to\infty} \frac{1}{2\pi i} \int\limits_{C_g} \exp(tp) p (p^2 + \omega^2)^{-1} \exp(-\sqrt{p}\,\alpha) dp$ gilt. Das mittels des Residuensatzes ermittelte Ergebnis ändert sich nicht, wenn man $r \to +0$ streben läßt. Bei diesem Grenzprozeß bleibt der Integrand im letzten Integral beschränkt und die Länge des Kreises C_r strebt gegen null. Beachtet man, daß für die Funktionswerte von $\sqrt{p}$ in jedem Punkt des oberen Ufers der Hauptwert zu nehmen ist, während es im entsprechenden Punkt des unteren Ufers der mit -1 multiplizierte Hauptwert ist, so ergibt sich in der obigen Formel für $- \lim\limits_{R\to\infty} \frac{1}{2\pi i} \int\limits_{C_g} (\ldots) dp$ schließlich

$$-(1/\pi) \int\limits_0^\infty \exp(-tx)\, x (x^2 + \omega^2)^{-1} \sin(\alpha\sqrt{x}) dx.$$

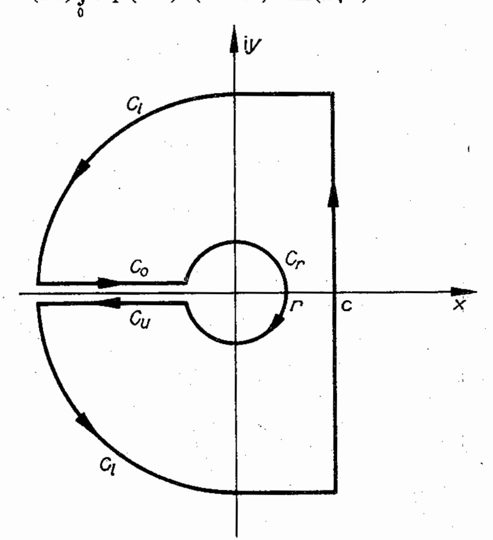


Abb. 13.4.7–2 Integrationsweg bei Bsp. 13.4.7–5

13.4.7.3. Cauchyscher Integralsatz. Er kann als Spezialfall des Residuensatzes, wobei ∞ nicht beteiligt ist, aufgefaßt werden. Ist nämlich $w(z)$ in G_1 überall holomorph (vgl. 13.4.7.2.), so ist $\oint_C w(z)\,dz = 0$. Eine andere Formulierung des *Cauchyschen Integralsatzes* lautet: Die beiden Integrale $\int_{C_1} w(z)\,dz$ und $\int_{C_2} w(z)\,dz$, wobei C_1 und C_2 sowohl einen gemeinsamen Anfangspunkt als auch einen gemeinsamen Endpunkt besitzen, sind einander gleich, falls es einen stetigen Deformationsprozeß gibt, der C_2 in C_1 überführt, wobei stets nur Punkte überstrichen werden, in denen $w(z)$ holomorph ist. In diesem Fall ist $\int_a^z w(t)\,dt$ bei festem a nicht vom Integrationsweg abhängig, sondern nur eine Funktion $W(z)$ der oberen Integrationsgrenze, und es gilt $W'(z) = w(z)$. $W(z)$ heißt Stammfunktion von $w(z)$. Sind in diesem Fall $z = a$ und $z = b$ Anfangs- bzw. Endpunkt von C, so gilt $\int_C w(z)\,dt = W(b) - W(a)$. Alle Lösungen $W(z)$ von $W'(z) = w(z)$ unterscheiden sich jeweils nur um eine additive Konstante. Man spricht in diesem Zusammenhang vom *unbestimmten Integral* $W(z) = \int w(z)\,dz$.

13.4.7.4. Unbestimmte Integrale einiger elementarer Funktionen. Die aus Abschnitt 13.3.4.6. folgenden Differentiationsformeln $(e^z)' = e^z$, $(\cos z)' = -\sin z$, $(\sin z)' = \cos z$, $(\tan z)' = (\cos z)^{-2} = 1 + (\tan z)^2$, $(\cot z)' = -(\sin z)^{-2} = -(1 + (\cot z)^2)$, $(\cosh z)' = \sinh z$, $(\sinh z)' = \cosh z$, $(\tanh z)' = (\cosh z)^{-2} = 1 - (\tanh z)^2$, $(\coth z)' = -(\sinh z)^{-2} = 1 - (\coth z)^2$, $(\ln z)' = 1/z$ ($z \neq 0$, z nicht negativ reell), $(z^\alpha)' = \alpha z^{\alpha-1}$ (Hauptwerte der Potenzen, $z \neq 0$, z nicht negativ-reell), $(\arcsin z)' = (1 - z^2)^{-1/2}$ (Hauptwert der Wurzel im Fall $|z| < 1$), $(\arccos z)' = -(1 - z^2)^{-1/2}$ (Hauptwert der Wurzel im Fall $|z| < 1$), $(\arctan z)' = (1 + z^2)^{-1}$, $(\operatorname{arccot} z)' = -(1 + z^2)^{-1}$, $(\operatorname{arcosh} z)' = (z^2 - 1)^{-1/2}$ (Hauptwert der Wurzel im Fall $|z| > 1$), $(\operatorname{arsinh} z)' = (1 + z^2)^{-1/2}$ (Hauptwert der Wurzel im Fall $|z| < 1$), $(\operatorname{artanh} z)' = (1 - z^2)^{-1}$, $(\operatorname{arcoth} z)' = (1 - z^2)^{-1}$ können gemäß 13.4.7.3. als unbestimmte Integrale umgeschrieben werden: $\int e^z dz = e^z$, $\int \sin z\,dz = -\cos z$ usw.

Riemannsche Fläche des komplexen Logarithmus. Man schneide die Gaußsche Zahlenebene längs der negativ-reellen Achse auf. Es entsteht dort ein *oberes* und ein *unteres Ufer*. Unendlich viele solcher aufgeschnittenen Exemplare – man spricht vom *Blatt* k ($k = 0, \pm 1, \pm 2, \ldots$) – denkt man sich übereinandergelegt. *Identifiziert* man nunmehr in diesem Stapel das obere Ufer von Blatt k mit dem unteren Ufer von Blatt $k + 1$ ($k = 0, \pm 1, \pm 2, \ldots$), so entsteht die *Logarithmusfläche* mit den beiden *logarithmischen Windungspunkten* an den Stellen 0 und ∞. Diese *Riemannsche Fläche* ist Definitionsbereich des komplexen Logarithmus, insbesondere ist das Blatt k ($k = 0, \pm 1, \ldots$) jeweils Definitionsbereich des k-ten Zweiges der Logarithmusfunktion. Während beispielsweise der Hauptwert des Logarithmus $\ln z$ (vgl. 13.3.4.6.) längs der negativ-reellen Achse unstetig ist – der Sprung beträgt dort stets $2\pi i$ – ist der komplexe Logarithmus dort stetig und sogar holomorph – auch dort ist die Ableitung des komplexen Logarithmus gleich $(1/z)$ – denn er ist mit der positiv-rellen Zahl a in der Umgebung der Stelle $z = -a$, die im oberen Ufer des Blattes k (das ist gleich der Stelle $z = -a$ im unteren Ufer des Blattes $k + 1$) liegt, gleich $\ln z + 2k\pi i$ im Fall $\operatorname{In} z \geqq 0$ und gleich $\ln z + (2k + 1)\pi i$ im Fall $\operatorname{In} z < 0$. Damit sind die Formeln $(\ln z)' = 1/z$ und $\int (1/z)\,dz = \ln z$ auch für negativ-reelles z gültig, falls man unter $\ln z$ den komplexen Logarithmus versteht und z auf der Riemannschen Fläche des Logarithmus variieren läßt. Diese Bemerkung ist sinngemäß auf $z^\alpha = e^{\alpha \ln z}$ zu übertragen.

Partialbruchzerlegung. Um eine *rationale Funktion* $R(z) = P(z)/Q(z)$ (P, Q = Polynome, der Grad von P sei gleich m, der von Q gleich n) in Partialbrüche zu zerlegen, stellt man zunächst die Produktdarstellung des Nenners her (vgl. 13.2.4.1.): $Q(z) = a_n(z - z_1)^{l_1}(z - z_2)^{l_2} \cdot \ldots \cdot (z - z_r)^{l_r}$ $(l_1 + l_2 \ldots + l_r = n)$. Dann gilt $R(z) = \left(\sum_{\varrho=1}^{r} \sum_{\lambda=1}^{l_\varrho} A_{\varrho\lambda}(z - z_\varrho)^{-\lambda}\right) + H(z)$ mit $H(z) \equiv 0$, falls $m < n$ und $H(z) = D_0 + \ldots + D_{m-n} z^{m-n}$, falls $m - n \geqq 0$ ist. Zur Berechnung der Konstanten $A_{\varrho\lambda}$ und D_ν ($\nu = 0, \ldots, m - n$) multipliziert man beide Seiten des Ansatzes für $R(z)$ mit dem Nenner $Q(z)$ und setzt danach für z speziell $z_1, \ldots, z_r$ ein. Zur Bestimmung der restlichen Konstanten von $A_{\varrho\lambda}$ und D_ν ordnet man in der erhaltenen Gleichung beide Seiten nach Potenzen von z und führt anschließend einen *Koeffizientenvergleich* durch. Sind die Koeffizienten der Polynome P und Q alle reell und soll die *Partialbruchzerlegung im Reellen* durchgeführt werden, so wird man in der Produktdarstellung von $Q(z)$ je zwei Faktoren der Gestalt $(z - z_0)^l$ und $(z - z_0^*)^l$ (z_0 = nicht-reell) zu $z^2 + pz + q$ zusammenfassen mit den *reellen* Koeffizienten $p = -(z_0 + z_0^*)$ und $q = z_0 z_0^* = |z_0|^2$. Dieser Faktor gibt dann in der Partialbruchzerlegung von $R(z)$ zu einer Summe der Gestalt $\sum_{\lambda=1}^{l} (B_\lambda z + C_\lambda)(z^2 + pz + q)^{-\lambda}$ Anlaß mit den noch zu bestimmenden *reellen* Konstanten B_λ und C_λ.

Unbestimmte Integration rationaler Funktionen. Infolge der Partialbruchzerlegung genügt die Angabe der Integration der Partialbrüche. Bleibt man im Komplexen, so genügt die Angabe von

$$\int z^n dz = z^{n+1}/(n+1) \quad (n = 0, 1, 2, \ldots),$$

$$\int (z - z_0)^{-n} dz = (z - z_0)^{-n+1}/(1 - n) \quad (n = 2, 3, \ldots),$$

$$\int (z - z_0)^{-1} dz = \ln(z - z_0).$$

Soll jedoch $\int R(x)\,dx$ mit $R(x) = P(x)/Q(x)$ (P, Q haben reelle Koeffizienten, x = reell) ganz im Reellen ausgewertet werden, so ist anzugeben:

$$\int x^n dx = x^{n+1}/(n+1) \quad (n = 0, 1, 2, \ldots),$$

$$\int (x - x_0)^{-1} dx = \ln|x - x_0|,$$

$$J_1 = \int (x^2 + px + q)^{-1} dx = 2(4q - p^2)^{-1/2} \arctan[(2x + p)(4q - p^2)^{-1/2}] \quad (4q - p^2 > 0),$$

$$\int (Bx + C)(x^2 + px + q)^{-1} dx = (B/2)\ln(x^2 + px + q) + (C - (pB)/2)J_1,$$

$$J_l = \int (x^2 + px + q)^{-l} dx = (l - 1)^{-1}(4q - p^2)^{-1}(2x + p)(x^2 + px + q)^{-(l-1)} + (4l - 6)(l - 1)^{-1}(4q - p^2)^{-1}J_{l-1} \quad (l = 2, 3, \ldots),$$

$$\int (Bx + C)(x^2 + px + q)^{-l} dx = -(B/2)(l - 1)^{-1}(x^2 + px + q)^{-(l-1)} + (C - (pB)/2)J_l \quad (l = 2, \ldots).$$

Unbestimmte Integration weiterer Funktionenklassen. Im Folgenden bedeutet $R(u, v)$ eine rationale Funktion von u und v, d.h. $R(u, v)$ ist als Quotient von $P(u, v)$ und $Q(u, v)$

darstellbar, wobei sowohl P als auch Q sowohl bezüglich u als auch bezüglich v Polynome sind.

$\int R(z,[(az+b)/(cz+d)]^{1/n})\mathrm{d}z$ $(ad-bc \neq 0)$. Durch die Substitution $t=[(az+b)/(cz+d)]^{1/n}$ (vgl. 13.4.2.3.) ergibt sich bezüglich t ein Integral über eine rationale Funktion von t.

$\int R(\mathrm{e}^z)\mathrm{d}z$. Durch die Substitution $t=\mathrm{e}^z$ ergibt sich ein Integral über eine rationale Funktion von t.

$\int R(x,(ax^2+2bx+c)^{1/2})\mathrm{d}x$ $(a \neq 0,\ a,b,c,x$ reell). 1. Fall: $ac-b^2>0$: Die Substitution $t=(ax+b)(ac-b^2)^{-1/2}$ führt auf $\int R_1(t,(t^2+1)^{1/2})\mathrm{d}t$. Die weitere Substitution $t=\sinh u$ führt auf $\int R_2(\mathrm{e}^u)\mathrm{d}u$. 2. Fall: $ac-b^2<0$: Die Substitution $t=(ax+b)(b^2-ac)^{-1/2}$ führt auf $\int R_1[t,(t^2-1)^{1/2}]\mathrm{d}t$, falls $a>0$ und auf $\int R_2(t,(1-t^2)^{1/2})\mathrm{d}t$, falls $a<0$ ist. Im ersten Fall substituiert man weiter $t=\cosh u$, im zweiten $t=\cos u$ und erhält $\int R_3(\mathrm{e}^u)\mathrm{d}u$ bzw. $\int R_4(\cos u, \sin u)\mathrm{d}u$. Im letzten Integral führt die Substitution $w=\tan(u/2)$ zu einem Integral über eine rationale Funktion von w.

Partielle Integration.

$\int u'(z)v(z)\mathrm{d}z = u(z)v(z) - \int u(z)v'(z)\mathrm{d}z$.

Integrale der Gestalt $\int R'(z)\ln z\,\mathrm{d}z$, $\int R'(z)\arctan z\,\mathrm{d}z$, $\int R'(z)\arcsin z\,\mathrm{d}z$, wobei $R'(z)$ die Ableitung einer rationalen Funktion bedeutet, lassen sich durch partielle Integration auf bereits angegebene Fälle zurückführen. Durch ein- oder mehrmalige partielle Integration lassen sich ferner Integrale der Gestalt $\int P(z)\mathrm{e}^{az}\mathrm{d}z$, $\int P(z)\sin(az+b)\mathrm{d}z$, $\int P(z)\cos(az+b)\mathrm{d}z$, $\int P(z)\mathrm{e}^{az}\sin(bz+c)\mathrm{d}z$ und $\int P(z)\mathrm{e}^{az}\cos(bz+c)\mathrm{d}z$ behandeln, wobei $P(z)$ ein Polynom von z bedeutet.

13.5. Gewöhnliche Differentialgleichungen

Unter einer *gewöhnlichen Differentialgleichung (Dgl.) n-ter Ordnung* für eine Funktion $y=y(x)$ versteht man eine Gleichung zwischen der unabhängigen Veränderlichen x, der abhängigen Veränderlichen y und den Ableitungen $y', y'', \dots, y^{(n)}$ für jeden Wert x des Definitionsbereiches von $y=y(x)$, wobei $y^{(n)}$ in der Gleichung tatsächlich vorkommt. Also kann eine gewöhnliche Dgl. in der Gestalt $F(x,y,y',\dots,y^{(n)})=0$ angegeben werden. $y=y(x)$ $(x \in D)$ heißt *Lösung* der Dgl., falls für alle $x \in D$ stets $F[x,y(x),y'(x),\dots,y^{(n)}(x)]=0$ gilt. Ist eine Auflösung der Dgl. nach $y^{(n)}$ möglich, also $y^{(n)}=f(x,y,y',\dots,y^{(n-1)})$, so liegt eine Dgl. in *expliziter Gestalt* vor.

13.5.1. Gewöhnliche lineare Differentialgleichungen

Unter einer *gewöhnlichen linearen Differentialgleichung n-ter Ordnung* für eine Funktion $y=y(x)$ versteht man $\sum_{\nu=0}^{n} a_\nu(x)y^{(\nu)}=g(x)$ mit $a_n(x) \not\equiv 0$ $(x \in D)$ – d.h. $a_n(x)$ ist in D nicht überall gleich null –, wobei D der gemeinsame Definitionsbereich der bekannten *Koeffizientenfunktionen* $a_\nu(x)$ $(\nu=0,\dots,n)$ und des bekannten *Störgliedes* $g(x)$ ist. Die Dgl. heißt darüber hinaus *inhomogen*, falls $g(x) \not\equiv 0$ $(x \in D)$, *homogen*, falls $g(x) \equiv 0$ $(x \in D)$ gilt. Wird die Bedingung $a_n(x) \not\equiv 0$ zu $a_n(x) \neq 0$ $(x \in D)$ – d.h. $a_n(x)$ ist in D nirgends gleich null – verschärft, so spricht man von einer *expliziten* Dgl., weil dann eine Auflösung nach $y^{(n)}$ möglich ist.

13.5.1.1. Lösungsstruktur linearer Differentialgleichungen. Analog zu 13.2.3.1. gilt: Sämtliche Lösungen der linearen Dgl. n-ter Ordnung werden in der *allgemeinen Lösung* angegeben. Diese ist gleich einer *partikulären* (d.h. speziellen) Lösung der gegebenen *inhomogenen* Dgl. plus der *allgemeinen* Lösung der zugehörigen *homogenen* Dgl. Sämtliche Lösungen der zugehörigen homogenen Dgl. bilden einen Vektorraum der Dimension n.

13.5.1.2. Lineare homogene Differentialgleichungen mit konstanten Koeffizienten. Zur Herstellung einer Basis des Lösungsraumes (vgl. 13.5.1.1.) der linearen homogenen Dgl. $\sum_{\nu=0}^{n} a_\nu y_h^{(\nu)}=0$. (Der Index h weist auf die *homogene* Dgl. hin) mit den *konstanten* Koeffizienten $a_0,\dots,a_n$ $(a_n \neq 0)$ benötigt man alle Lösungen λ_ϱ $(\varrho=1,\dots,r)$ mit den dazugehörigen Vielfachheiten l_ϱ $(\varrho=1,\dots,r)$ der *charakteristischen Gleichung* $\sum_{\nu=0}^{n} a_\nu\lambda^\nu=0$ (vgl. 13.2.4.1.). Eine – aus n Funktionen bestehende – Basis lautet $x^\mu\exp(\lambda_\varrho x)$ $(\varrho=1,\dots,r;$ $\mu=0,1,\dots,l_\varrho-1)$. Sind die Koeffizienten $a_0,\dots,a_n$ alle reell, so gibt es zu jeder nicht-reellen Lösung $\lambda=\alpha+\mathrm{i}\beta$ (mit der Vielfachheit l) der charakteristischen Gleichung die weitere Lösung $\lambda^*=\alpha-\mathrm{i}\beta$ mit der Vielfachheit l. Man kann in diesem Fall zu einer *reellen Basis* des Lösungsraumes übergehen, indem man die beiden Basiselemente $x^\mu\exp(\lambda x)$ und $x^\mu\exp(\lambda^* x)$ durch die beiden reellen Basiselemente

$(1/2)x^\mu[\exp(\lambda x)+\exp(\lambda^* x)] = x^\mu\exp(\alpha x)\cos(\beta x)$ und
$(1/2\mathrm{i})x^\mu[\exp(\lambda x)-\exp(\lambda^* x)] = x^\mu\exp(\alpha x)\sin(\beta x)$ ersetzt.

13.5.1.3. Partikuläre Lösung der linearen inhomogenen Differentialgleichungen mit konstanten Koeffizienten.

Ansatzmethode. Hat das Störglied $g(x)$ der Dgl. $\sum_{\nu=0}^{n} a_\nu y^{(\nu)} = g(x)$ $(a_\nu=\text{konst.};\ \nu=0,\dots,n;\ a_n \neq 0)$ die Struktur $g(x)=\sum_{\mu=0}^{m} b_\mu x^\mu \mathrm{e}^{qx}$ $(b_m \neq 0)$, so macht man für eine partikuläre Lösung $y_p(x)$ den Ansatz $y_p(x)=\sum_{\mu=0}^{m} B_\mu x^\mu \mathrm{e}^{qx} x^l$ mit $l=0$, falls q keine Lösung der charakteristischen Gleichung $\sum_{\nu=0}^{n} a_\nu\lambda^\nu=0$ ist; andernfalls soll l gleich der Vielfachheit der Lösung q der charakteristischen Gleichung sein. Zur Bestimmung von $B_0,\dots,B_m$ setzt man den Ansatz in die Dgl. ein, dividiert anschließend beide Seiten durch e^{qx}, ordnet danach nach Potenzen von x und führt schließlich einen Koeffizientenvergleich durch. Es ergibt sich ein lineares Gleichungssystem von $m+1$ Gleichungen für $B_0,\dots,B_m$, das genau eine Lösung hat.

Erweiterung der Ansatzmethode. Hat das Störglied $g(x)$ der Dgl. $\sum_{\nu=0}^{n} a_\nu y^{(\nu)}=g(x)$ $(a_\nu=\text{konst.};\ a_n \neq 0)$ die Gestalt

$g(x)=\sum_{\mu=0}^{m} b_\mu x^\mu \mathrm{e}^{\alpha x}\cos(\beta x)$ bzw. $g(x)=\sum_{\mu=0}^{m} b_\mu x^\mu \mathrm{e}^{\alpha x}\sin(\beta x)$

$(b_m \neq 0)$, wobei $a_0,\dots,a_n, b_0,\dots,b_m,\alpha,\beta$ als *reell* vorausgesetzt werden, so ist eine partikuläre Lösung $y_p(x)=\mathrm{Re}[Y_p(x)]$ bzw. $y_p(x)=\mathrm{Im}[Y_p(x)]$, wobei $Y_p(x)$ eine par-

tikuläre Lösung der Dgl. $\sum_{\nu=0}^{n} a_\nu Y^{(\nu)} = \sum_{\mu=0}^{m} b_\mu x^\mu e^{qx}$ mit $q = \alpha + i\beta$ ist.

Superponieren (auch bei *variablen* Koeffizienten gültig). Unterscheiden sich die s Dgl.en $\sum_{\nu=0}^{n} a_\nu y^{(\nu)} = g_\sigma(x)$ $(\sigma = 1,\ldots,s)$ nur in den Störgliedern und ist $y_\sigma(x)$ $(\sigma = 1,\ldots,s)$ jeweils eine partikuläre Lösung, so hat die Dgl. $\sum_{\nu=0}^{n} a_\nu y^{(\nu)} = \sum_{\sigma=1}^{s} c_\sigma g_\sigma(x)$ die partikuläre Lösung $y_p(x) = \sum_{\sigma=1}^{s} c_\sigma y_\sigma(x)$.

Variation der Konstanten (auch bei variablen Koeffizienten gültig). Ist von der zur Dgl. $\sum_{\nu=0}^{n} a_\nu y^{(\nu)} = g(x)$ $(a_n \neq 0)$ gehörigen homogenen Dgl. $\sum_{\nu=0}^{n} a_\nu y_h^{(\nu)} = 0$ eine – aus n Funktionen bestehende – Basis $y_{h_\nu}(x)$ $(\nu = 1,\ldots,n)$ des Lösungsraumes und damit die allgemeine Lösung $y_h(x) = \sum_{\nu=1}^{n} C_\nu y_{h_\nu}$ bekannt, so ergibt sich durch Ersetzen von C_ν $(\nu = 1,\ldots,n)$ durch noch zu bestimmende Funktionen $u_\nu(x)$ (diese Ersetzung heißt *Variation der Konstanten*) eine partikuläre Lösung $y_p(x) = \sum_{\nu=1}^{n} u_\nu(x) y_{h_\nu}(x)$ der inhomogenen Dgl., falls die n Funktionen $u_\nu'(x)$ $(\nu = 1,\ldots,n)$ dem aus n Gleichungen bestehenden linearen Gleichungssystem $\sum_{\nu=1}^{n} u_\nu'(x) y_{h_\nu}^{(\mu)}(x) = 0$ $(\mu = 0,\ldots,n-2)$, $\sum_{\nu=1}^{n} u_\nu'(x) y_{h_\nu}^{(n-1)} = a_n^{-1} g(x)$ genügen. Da nur ein y_p benötigt wird, kann bei der Bestimmung von $u(x)$ aus $u'(x)$ auf die Integrationskonstante verzichtet werden.

Bsp. 13.5.1–1. Greift an einem Teilchen der Masse m, das längs einer waagerechten x-Achse mit einer geschwindigkeitsproportionalen Reibungskraft beweglich ist, eine Feder an, deren anderer Randpunkt eine vorgeschriebene Bewegung ausführt, so genügt $x = x(t)$ (x = Ort des Teilchens; t = Zeit) der linearen Dgl. 2. Ordnung $m\ddot{x} + \alpha\dot{x} + kx = F(t)$. In einem elektrischen Schwingkreis gilt $L\ddot{q} + R\dot{q} + C^{-1}q = U(t)$ (Ladung auf der Kondensatorplatte q, Selbstinduktion L, Widerstand R, Kapazität C, Spannungsquelle $U(t)$). Beide genannten Dgl.en werden von der Dgl. $\ddot{y} + 2\delta\dot{y} + \omega_0^2 y = g(t)$ für $y = y(t)$ erfaßt, wobei nur der Fall $0 < \delta < \omega_0$ (*geringe Dämpfung*) vorgeführt werde. Die charakteristische Gleichung $\lambda^2 + 2\delta\lambda + \omega_0^2 = 0$ hat die Lösungen $\lambda_{1,2} = -\delta \pm i\omega(\omega = (\omega_0^2 - \delta^2)^{1/2})$. Folglich ist $e^{-\delta t}e^{i\omega t}$, $e^{-\delta t}e^{-i\omega t}$ und damit $e^{-\delta t}\cos(\omega t), e^{-\delta t}\sin(\omega t)$ eine Basis des Lösungsraumes von $\ddot{y}_h + 2\delta\dot{y}_h + \omega_0^2 y_h = 0$ (*Dgl. der freien gedämpften Schwingung*), also ist $y_h = e^{-\delta t}[C_1\cos(\omega t) + C_2\sin(\omega t)] = Ae^{-\delta t}\cos(\omega t - \varphi)$ mit $A = (C_1^2 + C_2^2)^{1/2}$, $C_1 = A\cos\varphi$ und $C_2 = A\sin\varphi$. Hat das Störglied die Periode T, d. h. gilt $g(t + kT) = g(t)$ $(-\infty < t < \infty;\ k = \pm 1, \pm 2, \ldots)$ und liegt die Fourier-Entwicklung (vgl. 13.1.4.1.)

$$g(t) = c_0/2 + \sum_{\nu=1}^{\infty} [c_\nu\cos(\omega_\nu t) + d_\nu\sin(\omega t)]$$

mit $\omega_\nu = (2\pi\nu)/T$ und den Fourier-Koeffizienten

$$c_\nu = (2/T)\int_{-T/2}^{T/2} g(t)\cos(\omega_\nu t)\,dt \quad (\nu = 0,1,\ldots),$$

$$d_\nu = (2/T)\int_{-T/2}^{T/2} g(t)\sin(\omega_\nu t)\,dt \quad (\nu = 1,2,\ldots)$$ vor, so hat eine partikuläre Lösung $y_p(t)$ der inhomogenen Dgl. die Gestalt

$$y_p(t) = c_0 y_{p0}(t) + \sum_{\nu=1}^{\infty} [c_\nu y_{p1\nu}(t) + d_\nu y_{p2\nu}(t)]$$ mit $\ddot{y}_{p0} + 2\delta\dot{y}_{p0} + \omega_0^2 y_{p0} = 1/2$, $\ddot{y}_{1\nu} + 2\delta\dot{y}_{p1\nu} + \omega_0^2 y_{p1\nu} = \cos(\omega_\nu t)$, $\ddot{y}_{p2\nu} + 2\delta\dot{y}_{p2\nu} + \omega_0^2 y_{p2\nu} = \sin(\omega_\nu t)$, also $y_{p1_\nu} = \mathrm{Re}[Y_{p\nu}(t)]$ und $y_{p2\nu} = \mathrm{Im}[Y_{p_\nu}(t)]$ mit $\ddot{Y}_{p_\nu} + 2\delta\dot{Y}_{p_\nu} + \omega_0^2 Y_{p_\nu} = \exp(i\omega_\nu t)$. Wegen $\delta > 0$ sind sowohl $q = 0$ als auch $q = i\omega_\nu$ keine Lösungen der charakteristischen Gleichung. Also können die Ansätze $y_{p0}(t) = B_{00}$ und $Y_{p\nu} = B_{0\nu}\exp(i\omega_\nu t)$ gemacht werden. Einsetzen in die zugehörigen Dgl.en ergibt $B_{00} = (2\omega_0^2)^{-1}$ und $B_{0\nu} = (-\omega_\nu^2 + 2\delta\omega_\nu i + \omega_0^2)^{-1} = |-\omega_\nu^2 + 2\delta\omega_\nu i + \omega_0^2|^{-1}\exp(-i\varphi_\nu)$ mit $\cot\varphi_\nu = (2\delta\omega_\nu)^{-1}(\omega_0^2 - \omega_\nu^2)$ $(0 < \varphi_\nu < \pi)$ und damit $y_{p0} = (2\omega_0^2)^{-1}$, $Y_{p\nu} = [(\omega_0^2 - \omega_\nu^2)^2 + 4\delta^2\omega_\nu^2]^{-1/2}\exp[i(\omega_\nu t - \varphi_\nu)]$; also ist schließlich

$$y_p(t) = c_0(2\omega_0^2)^{-1} + \sum_{\nu=1}^{\infty} [(\omega_0^2 - \omega_\nu^2)^2 + 4\delta^2\omega_\nu^2]^{-1/2}[c_\nu\cos(\omega_\nu t - \varphi_\nu) + d_\nu\sin(\omega_\nu t - \varphi_\nu)] = c_0(2\omega_0^2)^{-1} + \sum_{\nu=1}^{\infty} [(\omega_0^2 - \omega_\nu^2)^2 + 4\delta^2\omega_\nu^2]E\chi 1 - 1/2 D_\nu\cos(\omega_\nu t - \psi_\nu)$$

(vgl. die analoge Umformung bei y_h). Ist $g(t)$ nicht periodisch, so kann nach einem Grenzübergang $T \to \infty$ die obige Untersuchung übertragen werden, wobei die Rolle der Fourier-Reihe vom *Fourier-Integral* (vgl. Bsp. 13.1.5–6) übernommen wird.

Zusatzbedingungen sind häufig Anfangsbedingungen, d. h. Vorgaben der gesuchten Funktion und deren Ableitungen an einer einzigen Stelle (z. B. Anfangslage, Anfangsgeschwindigkeit), oder *Randbedingungen*, d. h. solche Vorgaben an mehr als einer Stelle. Es wird damit die gegebene Dgl. zur *Anfangswertaufgabe* bzw. *Randwertaufgabe* ergänzt.

Bsp. 13.5.1–2. Für die Durchbiegung $w = w(x)$ eines Balkens, der sich längs der x-Achse erstreckt $(0 \leq x \leq l)$ und das Flächenträgheitsmoment $J(x)$ hat, gilt $(EJ(x)w'')'' = p(x)$ (Elastizitätsmodul E), wobei $p(x)$ die senkrecht zur Balkenachse wirkende Streckenlast angibt. Ist der Balken beiderseits eingespannt, so liegen die Randbedingungen $w(0) = 0$, $w(l) = 0$, $w'(0) = 0$, $w'(l) = 0$ vor. Setzt man EJ = konst. voraus, so lautet die charakteristische Gleichung $EJ\lambda^4 = 0$. Deren Lösung ist $\lambda_1 = 0$ mit der Vielfachheit $l_1 = 4$. Also ist $1, x, x^2, x^3$ eine Basis des Lösungsraumes von $EJw_h'''' = 0$ und damit die allgemeine Lösung $w_h = C_1 + C_2x + C_3x^2 + C_4x^3$. Gemäß der Methode der Variation der Konstanten ist $w_p(x) = u_1(x) + u_2(x)x + u_3(x)x^2 + u_4(x)x^3$ eine partikuläre Lösung der inhomogenen Dgl., falls das folgende Gleichungssystem für $u_1', \ldots, u_4'$ erfüllt ist: $u_1' + xu_2' + x^2u_3' + x^3u_4' = 0$, $u_2' + 2xu_3' + 3x^2u_4' = 0$, $2u_3' + 6xu_4' = 0$, $6u_4' = p(x)/EJ$. Durch Integration der ermittelten Funktionen $u_1', \ldots, u_4'$ kann man für eine Lösung $u_1(x), \ldots, u_4(x)$ angeben: $u_4(x) = (6EJ)^{-1}\int_0^x p(\tilde{x})\,d\tilde{x}$,

$$u_3(x) = -(2EJ)^{-1}\int_0^x \tilde{x}p(\tilde{x})\,d\tilde{x},\ u_2(x) = (2EJ)^{-1}\int_0^x x^2p(\tilde{x})\,d\tilde{x},$$

$u_1(x) = -(6EJ)^{-1}\int_0^x \tilde{x}^3p(\tilde{x})\,d\tilde{x}$. Damit folgt nach geeigneter Zusammenfassung $w_p(x) = (6EJ)^{-1}\int_0^x (x - \tilde{x})^3p(\tilde{x})\,d\tilde{x}$.

Einsetzen der allgemeinen Lösung $w(x) = w_h(x) + w_p(x)$ in die Randbedingungen führt zu dem folgenden linearen Gleichungssystem für $C_1, \ldots, C_4$: $C_1 = 0$, $C_2 = 0$, $l^2C_3 + l^3C_4 = -w_p(l)$, $2lC_3 + 3l^2C_4 = -w_p'(l)$. Die Lösung $C_1, \ldots, C_4$

wird in die allgemeine Lösung $w(x) = w_h(x) + w_p(x)$ eingesetzt. Es ergibt sich schließlich $w(x) = \int_0^l G(x,\tilde{x})p(\tilde{x})\,d\tilde{x}$ mit der zur gegebenen Randwertaufgabe gehörigen *Greenschen Funktion (Einflußfunktion)* $G(x,\tilde{x})$, wobei $G(\tilde{x},x) = G(x,\tilde{x})$ ist und $G(x,\tilde{x}) = (6l^3EJ)^{-1}(-2x^3\tilde{x}^3 + 3lx^3\tilde{x}^2 + 3lx^2\tilde{x}^3 - 6l^2x^2\tilde{x}^2 + 3l^3x\tilde{x}^2 - l^3x^3)$ $(0 \leqq \tilde{x} \leqq x \leqq l)$ gilt. Ist speziell die Streckenlast $p(x)$ gleich $p_0 = \text{konst.}$ für $|x - x_\nu| \leqq \varepsilon$ und gleich 0 sonst in $0 \leqq x \leqq l$, so erhält man $w(x) = p_0 \int_{x_\nu-\varepsilon}^{x_\nu+\varepsilon} G(x,\tilde{x})\,d\tilde{x}$. Die durch diese Streckenlast hervorgerufene Gesamtkraft ist gleich $p_0^2\varepsilon = F_\nu$. Macht man nun einen Grenzübergang derart, daß $\varepsilon \to +0$, $p_0 \to \infty$ gehen, jedoch F_ν dabei fest bleibt, so erhält man die Belastung des Balkens durch eine Einzelkraft F_ν an der Stelle x_ν, benutzt hierfür in der Dgl. die Schreibweise $EJw'''' = F_\nu\delta(x - x_\nu)$ (δ = *Delta-Distribution*, sie ist *keine* Funktion) und erhält als Durchbiegung $w(x) = \lim_{\varepsilon\to+0} F_\nu(2\varepsilon)^{-1} \int_{x_\nu-\varepsilon}^{x_\nu+\varepsilon} G(x,\tilde{x})\,d\tilde{x} = F_\nu G(x,x_\nu)$. Also kann $G(x,\tilde{x})$ als Durchbiegung des Balkens an der Stelle x gedeutet werden, wenn an der Stelle $\tilde{x}$ eine Einzelkraft mit dem Zahlenwert 1 vorliegt. Nach dem Superpositionsgesetz gilt $w(x) = \sum_{\nu=1}^{n} F_\nu G(x,x_\nu)$, falls an den Stellen $x_1,\dots,x_n$ jeweils Einzelkräfte $F_1,\dots,F_n$ wirken. Tritt neben Einzelkräften noch eine Streckenlast $p(x)$ auf, so kann die Durchbiegung durch das *Lebesgue-Stieltjes-Integral* (vgl. 13.4.2.2.) $w(x) = \int_0^l G(x,\tilde{x})\,dF(\tilde{x})$ angegeben werden mit $F(x_\nu + 0) - F(x_\nu - 0) = F_\nu$ $(\nu = 1,\dots,n)$ und $F'(x) = p(x)$ für $x \neq x_\nu$.

13.5.1.4. Eulersche Differentialgleichung. Sie hat die Gestalt $\sum_{\nu=0}^{n} c_\nu x^\nu y^{(\nu)} = g(x)$, ist also eine lineare Dgl. mit *speziellen variablen Koeffizienten* $a_\nu(x) = c_\nu x^\nu$ $(\nu = 0,1,\dots,n)$. Setzt man $y(x) = Y(t)$ mit $t = \ln x$, so ergibt sich eine lineare Dgl. mit *konstanten* Koeffizienten für $Y(t)$. Infolgedessen führt das Einsetzen von $y_h(x) = x^\lambda$ in die zur Eulerschen Dgl. gehörige lineare homogene Dgl. mit anschließender Division durch x^λ zur charakteristischen Gleichung. Eine Basis des Lösungsraumes der homogenen Eulerschen Dgl. ist $(\ln x)^\mu x^{\lambda_\varrho}$ $(\varrho = 1,\dots,r;\ \mu = 0,\dots,l_\varrho - 1$; vgl. 13.5.1.2.). Die Ansatzmethode aus 13.5.1.3. ist übertragbar, indem man dort x jetzt durch $\ln x$ ersetzt:

$$\sum_{\nu=0}^{n} c_\nu x^\nu y^{(\nu)} = g(x) \text{ mit } g(x) = \sum_{\mu=0}^{m} b_\mu (\ln x)^\mu x^q$$

hat eine partikuläre Lösung der Gestalt

$$y_p(x) = \sum_{\mu=0}^{m} B_\mu (\ln x)^\mu x^q (\ln x)^l.$$

Bsp. 13.5.1–3. Im Inneren einer Vollkugel vom Radius R werde im Volumenelement ΔV und in der Zeit Δt die konstante Wärmemenge $Q\Delta V\Delta t$ erzeugt. Für das zur Ausdehnung gehörige Verschiebungsvektorfeld $\boldsymbol{u}(\boldsymbol{r})$ gilt $\boldsymbol{u}(\boldsymbol{r}) = f(r)\boldsymbol{r}$ $(r = |\boldsymbol{r}|)$ mit $\lim_{r\to+0} \boldsymbol{u}(\boldsymbol{r}) = 0$ und $rf''(r) + 4f'(r) = (1+\nu)(1-\nu)^{-1}\gamma T'(r)$ (Querdehnungszahl ν mit $0 < \nu < 1/2$; lineare Wärmeausdehnungszahl $\gamma > 0$), wobei für das Temperaturfeld $T = T(r)$ die Dgl. $T''(r) + (2/r)T'(r) = -(Q/\alpha)$ und die Randbedingungen „$\lim_{r\to+0} T(r)$ existiert" und $-\alpha T'(R) = \beta T(R)$ (α, β konstante Wärmeleitungs- bzw. Übergangszahlen) gelten. Multiplikation mit r^2 verwandelt die Dgl. für T in eine Eulersche Dgl. Einsetzen von $T_h(r) = r^\lambda$ in $r^2T_h'' + 2rT_h' = 0$ führt zur charakteristischen Gleichung $\lambda(\lambda - 1) + 2\lambda = 0$ mit den Lösungen $\lambda_1 = 0$ (Vielfachheit $l_1 = 1$) und $\lambda_2 = -1$ (Vielfachheit $l_2 = 1$). Also ergibt sich die allgemeine Lösung $T_h(r) = C_1 + C_2r^{-1}$. Da $q = 2$ keine Lösung der charakteristischen Gleichung ist, darf der Ansatz $T_p(r) = B_0r^2$ gemacht werden. Er liefert nach Einsetzen in $r^2T'' + 2rT' = -(Q/\alpha)r^2$ für B_0 den Wert $-Q/6\alpha$. Das Einsetzen von $T(r) = C_1 + C_2r^{-1} - Q(6\alpha)^{-1}r^2$ in die beiden Randbedingungen für T liefert $C_2 = 0$ und $C_1 = Q[R(3\beta)^{-1} + (6\alpha)^{-1}R^2]$. Die Dgl. für $f(r)$ geht nach Multiplikation mit r in eine Eulersche Dgl. über. Da hier nur e i n e Randbedingung für f vorliegt – nämlich $\lim_{r\to+0} f(r) = 0$ – tritt in der Lösungsformel für f noch eine willkürliche Konstante auf, die gemäß einer weiteren Randbedingung – z. B. Wert des Druckes auf der Oberfläche, d. h. $\sigma_r(R) = -p$ (σ_r = Radialspannung) mit $\sigma_r(r) = (1-2\nu)^{-1}2G[(1+\nu)f(r) + (1-\nu)rf'(r) - (1+\nu)\gamma T(r)]$ (Schubmodul G) – bestimmt werden kann.

13.5.1.5. Lineare homogene Differentialgleichungen 2. Ordnung mit variablen Koeffizienten. Da der inhomogene Fall mittels der Variation der Konstanten (vgl. 13.5.1.3.) auf den homogenen Fall zurückführbar ist, genügt die Behandlung der linearen *homogenen* Dgl., wobei das Wesentliche bereits bei Dgl. 2. Ordnung erkennbar ist.

Potenzreihenentwicklung der Lösung. Wenn in der Dgl. $p_0(z)w'' + p_1(z)w' + p_2(z)w = 0$ für $w = w(z)$ die Quotienten $p_\nu(z)/p_0(z)$ $(\nu = 1,2)$ an der Stelle $z = z_0$ holomorph (vgl. 13.3.3.9.) sind, so ist auch die allgemeine Lösung $w = w(z)$ an der Stelle $z = z_0$ holomorph und kann damit dort in eine Potenzreihe $w(z) = \sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu$ entwickelt werden. Ihr Konvergenzradius ist mindestens gleich der Entfernung der Stelle z_0 zur nächsten *Singularität* – d. h. bis zu einer solchen Stelle, an der die Holomorphie verletzt ist – von $p_\nu(z)/p_0(z)$ $(\nu = 1,2)$.

Bsp. 13.5.1–4. Beim Vorliegen der *Hermiteschen Dgl.* $w'' - 2zw' + 2nw = 0$ führt an der Stelle $z = z_0 = 0$ der Potenzreihenansatz $w(z) = \sum_{\nu=0}^{\infty} c_\nu z^\nu$ nach Einsetzen in die Dgl., Ordnen nach Potenzen von z und anschließendem Koeffizientenvergleich zur *Rekursionsformel* $c_{\nu+2} = -2(n - \nu)[(\nu+2)(\nu+1)]^{-1}c_\nu$ $(\nu = 0,1,2,\dots)$. Folglich haben die beiden Basiselemente des Lösungsraumes die Gestalt $\sum_{\mu=0}^{\infty} c_{2\mu}z^{2\mu}$ mit $c_{2\mu+2} = -2(n - 2\mu)[(2\mu+2)(2\mu+1)]^{-1}c_{2\mu}$ $(\mu = 0,1,\dots;\ |z| < \infty)$ und $\sum_{\mu=0}^{\infty} c_{2\mu+1}z^{2\mu+1}$ mit $c_{2\mu+3} = -2[n - (2\mu+1)][(2\mu+3)(2\mu+2)]^{-1}c_{2\mu+1}$ $(\mu = 0,1,\dots;\ |z| < \infty)$, wobei man sich für c_0 und c_1 jeweils eine von null verschiedene Zahl eingesetzt denken muß. Im Fall $n = 2m$ $(m = 0,1,\dots)$ reduziert sich das erste Basiselement, im Fall $n = 2m + 1$ $(m = 0,1,\dots)$ das zweite Basiselement auf ein Polynom von n-ten Grad. Wählt man c_0 bzw. c_1 derart, daß in diesem Polynom der Koeffizient von z^m gleich 2^n ist, so ergibt sich jeweils das *Hermitesche Polynom* $H_n(x)$ (vgl. Bsp. 13.1.4–7).

Stelle der Bestimmtheit. $z = z_0$ heißt Stelle der Bestimmtheit der Dgl. $p_0(z)w'' + p_1(z)w' + p_2(z)w = 0$, falls $p_\nu(z)/p_0(z)$ an der Stelle $z = z_0$ einen Pol höchstens ν-ter Ordnung besitzen $(\nu = 1,2$; vgl. 13.3.4.5.).

Der Lösungsansatz für ein erstes Basiselement ***an der***

Stelle der Bestimmtheit $z = z_0$ lautet $w(z) = (z - z_0)^\alpha \sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu = \sum_{\nu=0}^{\infty} c_\nu(z - z_0)^{\nu+\alpha}$ $(c_0 \neq 0)$. Für den Konvergenzradius von $\sum_{\nu=0}^{\infty} c_\nu(z - z_0)^\nu$ kann die oben gemachte Aussage übertragen werden. Entwicklung von $p_0(z)$, $p_1(z)$, $p_2(z)$ an der Stelle $z = z_0$ in Potenzreihen, Einsetzen des Ansatzes für $w(z)$ in die Dgl., Dividieren durch $(z - z_0)^\alpha$, Ordnen nach Potenzen von $z - z_0$ und Beginn des Koeffizientenvergleichs führt zu einer quadratischen Gleichung für α, deren Lösungen α_1 und α_2 so bezeichnet werden, daß $\operatorname{Re}\alpha_1 \geqq \operatorname{Re}\alpha_2$ ist. Mit $\alpha = \alpha_1$ wird der Koeffizientenvergleich weitergeführt. Es ergeben sich hiermit Gleichungen, aus denen die c_ν ermittelt werden können.

Bsp. 13.5.1–5. Die Lösungen der *Besselschen Dgl.*

$$z^2w'' + zw' + (z^2 - n^2)w = 0$$

(wegen des Gliedes n^2 in der Dgl. kann ohne Beschränkung an Allgemeinheit $\operatorname{Re} n \geqq 0$ vorausgesetzt werden) heißen *Zylinderfunktionen mit dem Index n*. $z = 0$ ist eine Stelle der Bestimmtheit. Der Ansatz $w(z) = \sum_{\nu=0}^{\infty} c_\nu z^{\nu+\alpha}$ $(c_0 \neq 0)$ führt zu den Lösungen $\alpha_1 = n$ und $\alpha_2 = -n$ der quadratischen Gleichung für α. Fortführung des Koeffizientenvergleichs mit $\alpha_1 = n$ ergibt einerseits $c_1 = 0$ und andererseits die Rekursionsformel $c_\nu = -[\nu(\nu + 2n)]^{-1} c_{\nu-2}$ $(\nu = 2, 3, \ldots)$. Hieraus folgt, daß alle c_ν mit ungeradem ν gleich null sind. Für gerades ν erhält man $c_{2\mu} = -[4\mu(\mu + n)]^{-1} c_{2\mu-2}$ $(\mu = 1, 2, \ldots)$ und damit $c_{2\mu} = (-1)^\mu [4^\mu \mu!(\mu + n)!]^{-1} n! c_0$ $(\mu = 1, 2, \ldots)$. Das Ergebnis ist auch für Werte von n mit $\operatorname{Re} n \geqq 0$ gültig, die nicht ganz sind (vgl. Bsp. 13.3.4–3). Man erhält damit, wenn man speziell $c_0 = (2^n n!)^{-1}$ setzt, für den Lösungsraum der Besselschen Dgl. als ein erstes Basiselement die *Besselfunktion (Zylinderfunktion erster Art) mit dem Index n*:

$$J_n(x) = (z/2)^n \sum_{\mu=0}^{\infty} (-1)^\mu [2^{2\mu} \mu!(\mu + n)!]^{-1} z^{2\mu}.$$

Herstellung eines zweiten Basiselementes. Ist die Differenz $\alpha_1 - \alpha_2$ nicht ganzzahlig, so ergibt sich ein zweites Basiselement mittels des obigen Ansatzes mit $\alpha = \alpha_2$. Ist jedoch $\alpha_1 - \alpha_2$ ganzzahlig, so gilt mit dem ersten Basiselement $w_1(z) = (z - z_0)^{\alpha_1} \sum_{\nu=0}^{\infty} c_\nu(\alpha_1)(z - z_0)^\nu [c_0(\alpha_1) \neq 0]$ für ein zweites Basiselement $w_2(z) = w_1(z) \cdot C \int \{[w_1(z)]^{-2} \exp[-\int p_1(z)/p_0(z)\,dz]\}\,dz = A w_1(z) \ln(z - z_0) + (z - z_0)^{\alpha_2} \sum_{\nu=0}^{\infty} c_\nu(\alpha_2)(z - z_0)^\nu$. Die Bestimmung von $w_2(z)$ kann also bei bekanntem $w_1(z)$ durch Berechnung des angegebenen Integrals geschehen. Man kann aber auch den letzten Ausdruck der Gleichungskette als Ansatz auffassen, ihn in die Dgl. einsetzen und schließlich einen Koeffizientenvergleich durchführen. Um zu prüfen, ob $A = 0$ oder $A \neq 0$ ist, kann man entweder das obige Integral diskutieren oder A beim Einsetzen des Ansatzes als Parameter auffassen und bei der weiteren Rechnung mitnehmen. $c_0(\alpha_2) = 0$ ist nur gesichert, wenn $\alpha_1 \neq \alpha_2$ ist.

Bsp. 13.5.1–6. Ein weiteres Basiselement des Lösungsraumes der Besselschen Dgl. ist im Fall „$\alpha_1 - \alpha_2 = 2n$ nicht ganz", d. h. $n \neq 0$, 1/2, 1, 3/2, 5/2, 3, ... die Funktion $J_{-n}(z)$. Es zeigt sich, daß dies sogar im Fall $n = m + (1/2)$ $(m = 0, 1, 2, \ldots)$ – man sagt n sei jetzt ein *halbzahliger Index* – gilt, obwohl dann $\alpha_1 - \alpha_2 = 2n$ ganz ist. Anstatt im Fall $n \neq 0, 1, 2, \ldots$ $J_{-n}(z)$ als weiteres Basiselement zu nehmen, ist es auch üblich, hierfür die *Neumannsche Funktion (Zylinderfunktion zweiter Art) mit nichtganzem Index n*

$$N_n(z) = [\sin(n\pi)]^{-1} [\cos(n\pi) J_n(z) - J_{-n}(z)]$$

zu wählen. Bei ganzzahligem n ist $J_{-n}(z) = (-1)^n J_n(z)$ und damit $J_n(z)$ als weiteres Basiselement unbrauchbar. Auch die obige Formel für $N_n(z)$ versagt in diesem Fall; man kann jedoch $N_n(z) = \lim_{p \to n} N_p(z)$ $(n = 0, 1, 2, \ldots;$ $p \neq 0, 1, 2, \ldots)$ als weiteres Basiselement wählen. Schließlich spannen die beiden folgenden *Hankelschen Funktionen (Zylinderfunktionen dritter Art) mit dem Index n* den Lösungsraum der Besselschen Dgl. auf:

$$H_n^{(1)}(z) = J_n(z) + \mathrm{i}N_n(z) \quad \text{und} \quad H_n^{(2)}(z) = J_n(z) - \mathrm{i}N_n(z).$$

Unendlich ist Stelle der Bestimmtheit. Die Stelle $z = \infty$ heißt Stelle der Bestimmtheit der Dgl. $p_0(z)w'' + p_1(z)w' + p_2(z)w = 0$, falls mit $w(z) = u(t)$ und $t = 1/z$ für die Dgl. $\tilde{p}_0(t)u'' + \tilde{p}_1(t)u' + \tilde{p}_2(t)u = 0$ die Stelle $t = 0$ eine Stelle der Bestimmtheit ist.

Bsp. 13.5.1–7. Für die *Legendresche Dgl.*

$$(z^2 - 1)w'' + 2zw' - n(n + 1)w = 0$$

ist $z = \infty$ eine Stelle der Bestimmtheit, denn mit $w(z) = u(t)$ $(z = 1/t)$ ist $w'(z) = -t^2u'(t)$, $w''(z) = t^4u''(t) + 2t^3u'(t)$ und damit $(t^2 - t^4)u'' - 2t^3u' - n(n + 1)u = 0$.

Asymptotische Entwicklung im Unendlichen. Will man beispielsweise die Besselfunktion $J_n(x)$ für große reelle x-Werte berechnen, so ist das zwar durch Entwicklung an der Stelle $x = 0$ *theoretisch* möglich – der Konvergenzradius ist ja unendlich – jedoch ist dies wegen schlechter Konvergenzgeschwindigkeit zumindest nicht zu empfehlen. Eine Entwicklung an der Stelle Unendlich stößt zunächst auf Schwierigkeiten, da Unendlich für die Besselsche Dgl. keine Stelle der Bestimmtheit ist. Es zeigt sich, daß man für den Fall $n > 0$ die Formel

$$J_n(x) = (2/\pi x)^{1/2} \cos(x - (\pi/2)n - \pi/4) \sum_{\nu=0}^{\infty} \{(-1)^\nu [n + 2\nu - (1/2)]!([2\nu]![n - 2\nu - (1/2)]!)^{-1}(2x)^{-2\nu}\}$$
$$- \sin(x - (\pi/2)n - \pi/4) \sum_{\nu=0}^{\infty} \{(-1)^\nu [n + 2\nu + (1/2)]![(2\nu + 1)!(n - 2\nu - (3/2))!]^{-1}(2x)^{-(2\nu+1)}\}$$

im folgenden Sinn angeben kann: Ist n *halbzahlig* – also $n = m + (1/2)$, $m = 0, 1, 2, \ldots$ –, so brechen die unendlichen Reihen ab, d. h. sie sind Polynome in $1/x$. (Um dies einzusehen, benötigt man die Gleichung $1/p! = 0$ $(p = -1, -2, \ldots)$. Sie ergibt sich bei der *analytischen Fortsetzung* der holomorphen Funktion $1/z!$ (vgl. Bsp. 13.3.4–3), d. h. bei der (auf genau eine Weise) möglichen Erweiterung des Definitionsbereiches $\operatorname{Re} z > -1$ von $1/z!$ derart, daß die Holomorphie erhalten bleibt). Ist n nicht-halbzahlig, so *divergieren* die obigen Reihen $\sum_{\nu=0}^{\infty}(\ldots)$, d. h. ersetzt man $\sum_{\nu=0}^{\infty}(\ldots)$ durch $\sum_{\nu=0}^{k}(\ldots) + R_k(x)$, so gilt nicht $\lim_{k\to\infty} R_k(x) = 0$; ja es gilt hier sogar $\lim_{k\to\infty} |R_k(x)| = \infty$. Trotzdem sind diese Reihen zur numerischen Auswertung brauchbar, denn es sind *asymptotische Reihen*, d. h. $R_k(x)$ strebt für $x \to \infty$ (nicht $k \to \infty$, sondern für jedes *feste* k) nach null, und zwar

von höherer Ordnung als der letzte Summand $c_k(x)$ in der Näherung $\sum_{\nu=0}^{k}(\ldots)$, d. h. es gilt $\lim_{x\to\infty}[R_k(x)/c_k(x)] = 0$. Die Reihen sind im vorliegenden Fall alternierend; daher ist der absolute Betrag des Fehlers, der beim Abbrechen der Reihen jeweils entsteht, kleiner als der Betrag des ersten vernachlässigten Gliedes.

13.5.1.6. Eigenwertaufgaben bei gewöhnlichen linearen Differentialgleichungen n-ter Ordnung. Sie bestehen aus einer linearen homogenen Dgl. n-ter Ordnung (vgl. 13.5.1.) und – in der Regel n – linearen homogenen Randbedingungen (vgl. Bsp. 13.1.5–10), wobei eine Konstante λ (*Eigenwertparameter* genannt), deren Werte einer Menge von reellen oder auch komplexen Zahlen zu entnehmen sind, entweder in der Dgl. oder in den Randbedingungen oder sowohl in der Dgl. als auch in den Randbedingungen auftritt. Alle diejenigen λ-Werte, für die es nichttriviale Lösungen $y(x)$ (d.h. $y(x) \not\equiv 0$) gibt, heißen *Eigenwerte*, die zugehörigen nichttrivialen Lösungen $y(x)$ heißen *Eigenlösungen*. Nimmt man zu den Eigenlösungen, die zum Eigenwert $\lambda = \lambda_k$ gehören, noch die Lösung $y \equiv 0$ hinzu, so bilden sie einen Vektorraum, *Eigenraum* genannt. Die *Dimension* des Eigenraumes (vgl. 13.1.4.) heißt *Vielfachheit* des zugehörigen Eigenwertes λ_k. Kommt der Eigenwertparameter λ nur in der Dgl. und dort linear vor, hat also die Dgl. die Gestalt $L[y] = \lambda M[y]$, wobei $L[y]$ und $M[y]$ lineare *Differentialausdrücke* $L[y] = \sum_{\nu=0}^{n} a_\nu(x)y^{(\nu)}(x)$, $M[y] = \sum_{\mu=0}^{m} b_\mu(x)y^{(\mu)}(x)$ sind $(m < n)$, bilden ferner $L[y]$ und $M[y]$ zusammen mit den gegebenen linearen homogenen Randbedingungen jeweils *hermitesche Operatoren* (vgl. 13.1.5.10., 13.1.5.11.), so sind alle Eigenwerte *reell*. Ist darüber hinaus die Eigenwertverteilung diskret, so existiert ein alle Eigenräume aufspannendes *Orthogonalsystem von Eigenlösungen* $\varphi_\mu(x)$ nach der jede *Vergleichsfunktion*, d. h. jede n-mal differenzierbare Funktion $u(x)$, die alle Randbedingungen erfüllt, in eine *Fourier-Reihe* $u(x) = \sum_\mu c_\mu\varphi_\mu(x)$ entwickelt werden kann mit den *Fourier-Koeffizienten* $c_\mu = \langle\varphi_\mu\,|\,u\rangle\|\varphi_\mu\|^{-2}$ (vgl. 13.1.4.1.), wobei das Skalarprodukt gemäß $\langle u\,|\,v\rangle = \int_J (M[u])^* v\,dx$ zu bilden ist und J das zur Aufgabe gehörende x-Intervall angibt.

Bsp. 13.5.1–8. Die zeitlich veränderliche Durchbiegung $G(x, t)$ $(0 \leqq x \leqq l,\ 0 \leqq t < \infty)$ – G weist auf die *Greensche Funktion* hin, vgl. Bsp. 13.5.1–2 – eines gebetteten Balkens genügt der *partiellen Dgl.* (d. h. die gesuchte Funktion hängt von *mehr* als einer unabhängigen Variablen ab) $EJ\dfrac{\partial^4 G}{\partial x^4} + \mu\dfrac{\partial^2 G}{\partial t^2} + kG = \delta(x - vt)$, wobei die auf der rechten Seite stehende *δ-Distribution* von einer Einzelkraft (Zahlenwert 1) verursacht wird, die mit der konstanten Geschwindigkeit v über den Balken läuft. Als Randbedingungen liege die Forderung vor, daß die zweiten und dritten partiellen Ableitungen von G nach x an den Stellen $x = 0$ und $x = l$ verschwinden, während die Anfangsbedingungen vorschreiben, daß zum Zeitpunkt $t = 0$ sowohl G als auch $\partial G/\partial t$ gleich null sind. Das *Separationsverfahren (Produktansatz, Fourier-Methode)* diskutiert zunächst nur die zugehörige *homogene* partielle Dgl. $EJ\dfrac{\partial^4 G_h}{\partial x^4} + \mu\dfrac{\partial^2 G_h}{\partial t^2} + kG_h = 0$ und beachtet nur die Randbedingungen. Das bedeutet: Der Ansatz $G_h(x, t) = X(x)T(t)$ führt nach Einsetzen in die homogene partielle Dgl. zu $EJX^{(4)}T + \mu X\ddot{T} + kXT = 0$ und damit durch *Separation* (d.h. Umformung, wobei auf der einen Seite eine Funktion steht, die nur von x abhängt, während auf der anderen Seite eine Funktion steht, die nur von t abhängt) zu $X^{(4)}/X = (EJ)^{-1}(-\mu\ddot{T}/T - k)$. Diese Gleichung kann nur bestehen, wenn beide Seiten gleich einer Konstanten $\tilde{\lambda}$ sind. Die gegebenen Randbedingungen führen infolge des Separationsansatzes zu $X''(0) = X''(l) = X'''(0) = X'''(l) = 0$. Damit liegt für $X = X(x)$ eine Eigenwertaufgabe mit dem Eigenwertparameter $\tilde{\lambda}$ vor. Setzt man die allgemeine Lösung der Dgl. $X^{(4)} = \tilde{\lambda}X$ in die Randbedingungen ein, so ergibt sich ein lineares homogenes Gleichungssystem für die vier willkürlichen Konstanten der allgemeinen Lösung. Nichttriviale Lösungen $X(x)$ ergeben sich nur, wenn die Koeffizientendeterminante des Gleichungssystems gleich null ist (*Eigenwertgleichung*). Mit der Abkürzung $\tilde{\lambda} = \lambda^4$ lautet die Eigenwertgleichung $\cosh(\lambda l)\cos(\lambda l) = 1$. Aus einer Skizze der Funktionen $1/\cosh(\lambda l)$ und $\cos(\lambda l)$ entnimmt man Näherungswerte für die Eigenwerte, die durch ein Iterationsverfahren (vgl. Bsp. 13.3.2–2) noch verbessert werden können. Durch Einsetzen der gefundenen Werte für λ in das lineare homogene Gleichungssystem für die Konstanten der allgemeinen Lösung der Dgl. für X ergeben sich zugehörige Werte dieser Konstanten und damit zugehörige Eigenfunktionen. Es ist $\lambda = 0$ Eigenwert und eine orthogonale Basis des zugehörigen zweidimensionalen Eigenraumes ist $X_{00} \equiv 1$ und $X_{01} = x - (l/2)$. Für die weiteren Eigenwerte erhält man $\lambda_1 l = 4{,}7300\ldots$, $\lambda_2 l = 7{,}853\ldots$, $\lambda_3 l = 10{,}996\ldots$ und $\lambda_m l = (\pi/2) + m\pi$ $(m = 4, 5, \ldots)$ im Rahmen der mitgeführten Stellenzahl. Zu λ_m $(m = 1, 2, \ldots)$ ist der zugehörige Eigenraum jeweils eindimensional und als Basis kann $X_m(x) = [\sin(\lambda_m l) - \sinh(\lambda_m l)][\cosh(\lambda_m x) + \cos(\lambda_m x)] + [\cosh(\lambda_m l) - \cos(\lambda_m l)][\sinh(\lambda_m x) + \sin(\lambda_m x)]$ genommen werden. An dieser Stelle ist die Hilfestellung des Separationsverfahrens für die Lösung der ursprünglichen Aufgabe beendet. Man benutzt das gefundene vollständige Orthogonalsystem $X_{00}(x)$, $X_{01}(x)$, $X_m(x)$ $(m = 1, 2, 3, \ldots)$ zur Entwicklung der gesuchten Funktion $G(x, t)$ bezüglich x, setzt also $G(x, t) = T_{00}(t)X_{00}(x) + T_{01}(t)X_{01}(x) + \sum_{m=1}^{\infty} T_m(t)X_m(x)$. Infolge des bisherigen Vorgehens erfüllt diese Formel die Randbedingungen für G. Die Fourier-Koeffizienten – sie hängen hier von t ab – sind derart zu bestimmen, daß G darüber hinaus der gegebenen Dgl. und den Anfangsbedingungen genügt. Die Formeln für die Fourier-Koeffizienten $T_{00}(t)$, $T_{01}(t)$, $T_m(t)$ sind nicht unmittelbar auswertbar, da $G(x, t)$ unbekannt ist. Das weitere Vorgehen wird an $T_m(t) = \|X_m\|^{-2}\int_0^l X_m(x)G(x, t)\,dx$ erläutert. Für $\ddot{T}_m(t)$ ergibt sich bei Beachtung der Dgl. für G zunächst $\ddot{T}_m(t) = \|X_m\|^{-2}\int_0^l X_m(x)\,(1/\mu)\,\{\delta(x - vt) - EJ\partial^4 G/\partial x^4 - kG\}\,dx$. Durch viermalige partielle Integration folgt bei Beachtung der Randbedingungen für $G(x, t)$ und von $X_m^{(4)} = \lambda_m^4 X_m$ weiter $\int_0^l X_m(x)\,(\partial^4 G/\partial x^4)\,dx = \lambda_m^4\int_0^l X_m(x)G(x, t)\,dx = \lambda_m^4\|X_m\|^2 T_m(t)$. Dies führt schließlich zur gewöhnlichen linearen inhomogenen Dgl. mit konstanten Koeffizienten für $T_m(t)$:

$$\ddot{T}_m + (1/\mu)\,(EJ\lambda_m^4 + k)T_m = \|X_m\|^{-2}(1/\mu)\int_0^l X_m(x)\delta(x - vt)\,dx.$$

Analog zum Bsp. 13.5.1–2 ist das letzte Integral gleich $X_m(vt)$. Die Formeln für die Fourier-Koeffizienten zeigen, daß die Anfangsbedingungen für $G(x,t)$ die Anfangsbedingungen $T_{00}(0) = \dot{T}_{00}(0) = T_{01}(0) = \dot{T}_{01}(0) = T_m(0) = \dot{T}_m(0) = 0$ zur Folge haben. Gemäß 13.5.1.2. und 13.5.1.3. bestimmt man die allgemeinen Lösungen $T_{00}(t)$, $T_{01}(t)$, $T_m(t)$ und ermittelt die zunächst noch willkürlichen Konstanten aus den Anfangsbedingungen.

Bsp. 13.5.1–9. Bei der quantenmechanischen Behandlung des linearen harmonischen Oszillators ist für $y = y(x)$ die Eigenwertaufgabe $H[y] = Ey$ $(-\infty < x < \infty$; E = Eigenwertparameter [*Quantenzahlparameter*]) zu lösen mit dem *Hamilton-Operator* (in der Ortsdarstellung) $H = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + \frac{m}{2}\omega^2 x^2$ und der Forderung – sie ist eine lineare homogene Randbedingung – $\int_{-\infty}^{\infty} |y(x)|^2 dx$ soll existieren. Mit $q = (m\omega/\hbar)^{1/2}x$, $w(q) = y(x)$, $\lambda = 2E/(\hbar\omega)$ liegt also die Eigenwertaufgabe $w''(q) - q^2 w = -\lambda w$, $\int_{-\infty}^{\infty} |w(q)|^2 dq < \infty$ vor. Eine Potenzreihenentwicklung von $w(q)$ an der Stelle $q = 0$ ist möglich, aber es bestehen dann Schwierigkeiten, die Bedingung $\int_{-\infty}^{\infty} |w(q)|^2 dq < \infty$ zu erfüllen. Auch die Entwicklung an der Stelle $q = \infty$ stößt zunächst auf Schwierigkeiten, da $q = \infty$ keine Stelle der Bestimmtheit der Dgl. ist. Versucht man die Dgl. zunächst näherungsweise mit dem Ansatz $w(q) = \exp(cq^2)$ zu lösen, so wird man wegen $[\exp(cq^2)]'' - q^2\exp(cq^2) = [(4c^2-1)q^2 + 2c]\exp(cq^2)$ für $c = \pm 1/2$ ansetzen und im Hinblick auf die Randbedingung nur $c = -1/2$ wählen. Der nunmehr für das exakte $w(q)$ naheliegende Ansatz $w(q) = W(q)\exp[(-1/2)q^2]$ liefert für $W(q)$ die *Hermitesche Dgl.* (vgl. Bsp. 13.5.1–4) $W'' - 2qW' + (\lambda - 1)W = 0$. $W(q)$ wird gemäß Bsp. 13.5.1–4 bestimmt. Es zeigt sich, daß nur solche $W(q)$ zur Erfüllung der Randbedingungen brauchbar sind, deren Potenzreihenentwicklungen (Entwicklungsstelle $q = 0$) abbrechen, die also – abgesehen von konstanten Faktoren – gleich Hermiteschen Polynomen sind. Hiermit ergeben sich somit die Eigenwerte durch diese *Abbruchbedingung*. Man erhält $\lambda = 2n + 1$ $(n = 0,1,2,\ldots)$ und damit $E_n = \hbar\omega[n + (1/2)]$ $(n = 0,1,2,\ldots)$. Ein zugehöriges vollständiges Orthonormalsystem von Eigenfunktionen lautet $H_n(q)\exp(-q^2/2)$ mit $q = (m\omega/\hbar)^{1/2}x$ [$H_n(q)$ = Hermitesches Polynom], wobei die Orthogonalität relativ zum Skalarprodukt $\langle u|v\rangle = \int_{-\infty}^{\infty} (u(x))^* v(x) dx$ zu verstehen ist.

13.5.2. Gewöhnliche Differentialgleichungen erster Ordnung

Die betrachteten Dgl.en sind i. allg. nichtlinear. Wenn die im folgenden behandelten Spezialfälle nicht vorliegen, ist man in der Regel auf numerische Verfahren angewiesen (vgl. 13.3.2.2. und 13.5.2.4.).

13.5.2.1. Differentialgleichungen mit trennbaren Veränderlichen. Unter einer gewöhnlichen Dgl. mit *trennbaren Veränderlichen* für die Funktion $y = y(x)$ versteht man $y' = g(x)h(y)$. Ist $y = y_0$ eine Nullstelle von $h(y)$, gilt also $h(y_0) = 0$, so ist $y \equiv y_0$ eine Lösung. Im Fall $h(y) \neq 0$ behandle man in $dy/dx = g(x)h(y)$ den Differentialquotienten als Bruch und t r e n n e x und y gemäß $dy/h(y) = g(x)dx$. Bildet man auf beiden Seiten je ein unbestimmtes Integral – unter Verzicht auf Integrationskonstanten – und fügt man danach auf einer Seite additiv eine Konstante C hinzu, so ergibt sich schließlich für die gesuchten Lösungen $y = y(x)$ die implizite Darstellung (vgl. 13.3.5.4.) $F(x,y) = 0$ mit $F(x,y) = \int (1/h(y))dy - \int g(x)dx - C$.

13.5.2.2. Lineare Differentialgleichungen 1. Ordnung mit variablen Koeffizienten. Liegt die gewöhnliche lineare Dgl. 1. Ordnung $a_1(x)y' + a_0(x)y = g(x)$ $(a_1(x) \neq 0)$ vor (vgl. 13.5.1.), so wird man zunächst die allgemeine Lösung $y_h(x)$ der zugehörigen homogenen Dgl. $a_1(x)y_h' + a_0(x)y_h = 0$ durch die Methode der Trennung der Veränderlichen (vgl. 13.5.2.1.) herstellen: $y_h(x) = C\exp\left(-\int \{a_0(x)/a_1(x)\}dx\right)$. Eine partikuläre Lösung $y_p(x)$ der gegebenen Dgl. wird mittels der Methode der *Variation der Konstanten* (vgl. 13.5.1.3.) bestimmt:

$y_p(x) = u(x)\tilde{y}_h(x)$ mit $\tilde{y}_h(x) = \exp\left(-\int \{a_0(x)/a_1(x)\}dx\right)$ und $u' = g(x)/(a_1(x)\tilde{y}_h(x))$. Die allgemeine Lösung der gegebenen Dgl. ist $y(x) = y_h(x) + y_p(x)$ (vgl. 13.5.1.1.).

13.5.2.3. Exakte Differentialgleichungen. Unter $P(x,y)dx + Q(x,y)dy = 0$ versteht man die Aufgabe, bei bekannten Funktionen $P(x,y)$ und $Q(x,y)$ entweder Funktionen $y = y(x)$ zu bestimmen, die der Dgl. (Division der obigen Gleichung durch dx) $P(x,y) + Q(x,y)y' = 0$ genügen, oder aber auch die Aufgabe, Funktionen $x = x(y)$ zu ermitteln, die der Dgl. $P(x,y)x'(y) + Q(x,y) = 0$ genügen. Die gegebene Dgl. heißt *exakt*, falls es im Definitionsbereich von P und Q eine Funktion $U = U(x,y)$ mit $\partial U/\partial x = P$ und $\partial U/\partial y = Q$ gibt. Hieraus folgt, daß die *Integrabilitätsbedingung* $\partial P/\partial y = \partial Q/\partial x$ notwendig dafür ist, daß die Dgl. exakt ist. Diese Bedingung ist sogar *hinreichend*, falls der gemeinsame Definitionsbereich von $P(x,y)$ und $Q(x,y)$ *einfach-zusammenhängend* ist, d. h. wenn sich dort jede geschlossene Kurve stets auf einen Punkt zusammenziehen läßt ohne bei dem Deformationsprozeß den Definitionsbereich zu verlassen. Die Lösungen $y = y(x)$ bzw. $x = x(y)$ einer exakten Dgl. werden implizit durch $U(x,y) - C = 0$ dargestellt. Multipliziert man die *nicht-exakte* Dgl. $P(x,y)dx + Q(x,y)dy = 0$ mit einer Funktion $\mu(x,y) \neq 0$ und ist die sich hierdurch ergebende Dgl. $(\mu P)dx + (\mu Q)dy = 0$ exakt, so heißt $\mu = \mu(x,y)$ *integrierender Faktor (Eulerscher Multiplikator)* der Dgl. Meist versucht man mit speziellen Ansätzen für $\mu(x,y)$ zum Ziel zu kommen, beispielsweise $\mu = \mu(x)$, $\mu = \mu(y)$, $\mu = \mu(x+y)$, $\mu = \mu(xy)$, $\mu = \mu(y/x)$, $\mu = \mu(x^2 + y^2)$ usw. Ob diese Versuche gelingen, hängt von der Struktur von P und Q ab.

Bsp. 13.5.2–1. Wählt man bei einem reversiblen thermodynamischen Prozeß die absolute Temperatur T und den Druck p als unabhängige Veränderliche, so nimmt bei einer (differentiell kleinen) Änderung der Temperatur T um dT und des Druckes p um dp das System die Wärmemenge $c_p(T,p)dT + \lambda_T(T,p)dp$ auf. Hierbei bedeuten $c_p(T,p)$ die spezifische Wärmekapazität bei konstantem Druck und $\lambda_T(T,p)$ die spezifische latente Wärme bei konstanter Temperatur. Faßt man den Ausdruck für die aufgenommene Wärmemenge als linke Seite einer Dgl. $c_p(T,p)dT + \lambda_T(T,p)dp = 0$ auf, so ist für diese Dgl. T ein integrierender Nenner, d. h. $1/T$ ein integrierender Faktor (vgl. *Zweiter Hauptsatz der Wärmelehre*, 2.6.3.).

Bsp. 13.5.2–2. In Fortführung des Bsp. 13.5.2–1 gibt $c_p(T,p)dT + \lambda_T(T,p)dp - p\left(\frac{\partial V}{\partial T}dT + \frac{\partial V}{\partial p}dp\right)$ (V = spezifi-

sches Volumen, $V = V(T,p)$ = Zustandsgleichung) die Änderung der inneren Energie an. Faßt man diesen Ausdruck als linke Seite einer Dgl. $c_p \mathrm{d}T + \lambda_T \mathrm{d}p - p\left(\frac{\partial V}{\partial T}\mathrm{d}T + \frac{\partial V}{\partial p}\mathrm{d}p\right) = 0$ auf, so ist diese Dgl. exakt (vgl. *Erster Hauptsatz der Wärmelehre*, 2.6.2.).

13.5.2.4. Das Verfahren von Runge-Kutta. Zur näherungsweisen Berechnung der Lösung $y = y(x)$ der Anfangswertaufgabe $y' = f(x,y)$, $y(x_0) = y_0$ und damit der nichtlinearen Integralgleichung $y(x) = y_0 + \int_{x_0}^{x} f[t, y(t)]\,\mathrm{d}t$ (vgl. Bsp. 13.3.1–1) an den Stellen $x_\nu = x_0 + \nu h$ ($\nu = 1,2,3,\ldots$; h = *Schrittweite*), d. h. zur Angabe von Näherungswerten y_ν für die exakten Werte $y(x_\nu) = y(x_{\nu-1}) + \int_{x_{\nu-1}}^{x_\nu} f[t, y(t)]\mathrm{d}t$ ($\nu = 1,2,\ldots$) werden zunächst vier vorläufige Näherungen y_ν^{I}, y_ν^{II}, y_ν^{III}, y_ν^{IV} für $y(x_\nu)$ bei bekannter Näherung $y_{\nu-1}$ ($\nu = 1,2,\ldots$) hergestellt, indem der Integrand jeweils durch eine konstante Funktion ersetzt wird. Der Wert des Integranden an der unteren Integrationsgrenze führt zu $y_\nu^{\mathrm{I}} = y_{\nu-1} + f(x_{\nu-1}, y_{\nu-1})h$, der Wert von $f(x,y)$ im Halbierungspunkt der Strecke zwischen $(x_{\nu-1}, y_{\nu-1})$ und $(x_\nu, y_\nu^{\mathrm{I}})$ liefert $y_\nu^{\mathrm{II}} = y_{\nu-1} + f[x_{\nu-1} + (h/2), (1/2)(y_{\nu-1} + y_\nu^{\mathrm{I}})]h$; analog wird $y_\nu^{\mathrm{III}} = y_{\nu-1} + f[x_{\nu-1} + (h/2), (1/2)(y_{\nu-1} + y_\nu^{\mathrm{II}})]h$ gebildet und schließlich mit $f(x,y)$ an der Stelle $(x_\nu, y_\nu^{\mathrm{III}})$ die Näherung $y_\nu^{\mathrm{IV}} = y_{\nu-1} + f(x_\nu, y_\nu^{\mathrm{III}})h$ gewonnen. Die Linearkombination $c_1 y_\nu^{\mathrm{I}} + c_2 y_\nu^{\mathrm{II}} + c_3 y_\nu^{\mathrm{III}} + c_4 y_\nu^{\mathrm{IV}}$ soll durch geeignete Wahl von $c_1, \ldots, c_4$ eine möglichst gute Näherung y_ν für $y(x_\nu)$ ergeben. Hierzu wird die Lösung $y(x)$ an der Stelle $x = x_{\nu-1}$ nach Taylor entwickelt und es werden die Taylor-Koeffizienten durch $f(x,y)$ und deren partielle Ableitungen an der Stelle $(x_{\nu-1}, y_{\nu-1})$ ausgedrückt. Setzt man danach $x = x_\nu$, so ergibt sich für $y(x_\nu)$ eine Potenzreihenentwicklung nach Potenzen von h. Nunmehr wird auch y_ν nach Potenzen von h entwickelt. Die Konstanten $c_1, \ldots, c_4$ werden derart bestimmt, daß beide Entwicklungen bis zu möglichst hohen Potenzen von h übereinstimmen (*Taylor-Abgleich*). Es ergibt sich $c_1 = 1/6$, $c_2 = 1/3$, $c_3 = 1/3$, $c_4 = 1/6$ und eine Übereinstimmung bis zum Glied mit h^4. Also ist schließlich
$y_{\nu+1} = y_\nu + k$ mit $k = (1/6)(k_1 + 2k_2 + 2k_3 + k_4)$ und
$k_1 = f(x_\nu, y_\nu)h, k_2 = f[x_\nu + (h/2), y_\nu + (k_1/2)]h$,
$k_3 = f[x_\nu + (h/2), y_\nu + (k_2/2)]h$, $k_4 = f[x_{\nu+1}, y_\nu + k_3]h$
($\nu = 0,1,2,\ldots$). Zur näherungsweisen Berechnung der Fehler der Näherungswerte $y_2, y_4, y_6, \ldots$ kann man $(1/15)(y_{2k} - y_{2k,G})$ ($k = 1,2,3,\ldots$) angeben, wobei $y_{2k,G}$ derjenige Näherungswert für $y(x_{2k})$ ist, der sich aus dem mit der *doppelten* Schrittweite $2h$ durchgeführten Runge-Kutta-Verfahren ergibt.

13.5.3. Gewöhnliche Differentialgleichungen 2. Ordnung

Das Verfahren von Picard-Lindelöf (vgl. 13.3.2.2. in Verbindung mit Bsp. 13.3.1–1) ist übertragbar. Für die Lösung $y = y(x)$ von $y'' = f(x,y,y')$, $y(x_0) = y_0, y'(x_0) = y_0'$ gilt – bei analogen Voraussetzungen wie in 13.3.2.2. – $y(x) = \lim_{k\to\infty} y_k(x)$ ($|x - x_0| < r$, r hinreichend klein), wobei die Funktionenfolge $y_0(x), y_1(x), \ldots$ unter Hinzuziehen der Hilfsfunktionenfolge $z_0(x), z_1(x), \ldots$ gemäß
$y_0(x) \equiv y_0$, $z_0(x) \equiv y_0'$, $y_k(x) = y_0 + \int_{x_0}^{x} z_{k-1}(t)\,\mathrm{d}t$, $z_k(x) = y_0' + \int_{x_0}^{x} f[t, y_{k-1}(t), z_{k-1}(t)]\mathrm{d}t$ ($k = 1,2,\ldots$) zu berechnen ist.
Auch das Verfahren von Runge-Kutta (vgl. 13.5.2.4.) ist übertragbar. Bei vorliegender Anfangswertaufgabe $y'' = f(x,y,y')$, $y(x_0) = y_0$, $y'(x_0) = y_0'$ stelle man das hierzu äquivalente Dgl.system $\boldsymbol{y}' = \boldsymbol{f}(x,\boldsymbol{y})$, $\boldsymbol{y}(x_0) = \boldsymbol{y}_0$ mit den einspaltigen Matrizen $\boldsymbol{y} = \begin{pmatrix} y \\ y' \end{pmatrix}$, $\boldsymbol{y}' = \begin{pmatrix} y' \\ y'' \end{pmatrix}$, $\boldsymbol{f}(x,\boldsymbol{y}) = \begin{pmatrix} y' \\ f(x,y) \end{pmatrix}$, $\boldsymbol{y}_0 = \begin{pmatrix} y_0 \\ y_0' \end{pmatrix}$ her und führe das Runge-Kutta-Verfahren durch, indem man dort $f(x,y)$, $y(x_\nu)$, y_ν, $k_1, \ldots, k_4$, k jetzt durch $\boldsymbol{f}(x,\boldsymbol{y})$, $\boldsymbol{y}(x_\nu)$, $\boldsymbol{y}_\nu$, $\boldsymbol{k}_1, \ldots, \boldsymbol{k}_4$, $\boldsymbol{k}$ ersetzt.
Im Folgenden werden solche Spezialfälle der Dgl. $y'' = f(x,y,y')$ betrachtet, deren Lösungen $y = y(x)$ oder deren Umkehrfunktionen $x = x(y)$ durch elementare Funktionen oder durch Integrale darstellbar sind oder sich aus Lösungen von Dgl. erster Ordnung ergeben.

13.5.3.1. Die Differentialgleichung $y'' = f(x)$. Ist $f(x)$ stückweise stetig, so kann die allgemeine Lösung der Dgl. $y'' = f(x)$ in der Gestalt $y(x) = \int_{x_0}^{x} (x - t)f(t)\mathrm{d}t + C_1(x - x_0) + C_2$ angegeben werden, wobei x_0 eine feste Zahl aus dem Definitionsbereich von f ist.

13.5.3.2. Die Differentialgleichung $y'' = f(y)$, Energiemethode. Die Dgl. $y'' = f(y)$ ist im Fall $y' \neq 0$ gleichbedeutend mit $\mathrm{d}[(1/2)y'^2]/\mathrm{d}x = f(y)y'$ [$(1/2)y'^2$ = *kinetische Energie* eines Teilchens, wobei der Zahlenwert der Masse gleich 1 ist, x als Zeit und y als Ort gedeutet werden] und damit $(1/2)y'^2 = \int f(y(x))y'(x)\mathrm{d}x + C$. Also kann wegen $y' = \pm\left(2\int f(y)\mathrm{d}y + 2C\right)^{1/2}$ die Lösung $y = y(x)$ durch Trennung der Veränderlichen (vgl. 13.5.2.1.) ermittelt werden.

Bsp. 13.5.3–1. Der Ausschlagwinkel φ des mathematischen Pendels als Funktion der Zeit t genügt der Dgl. $\ddot{\varphi} = -(g/l)\sin\varphi$ (g = Erdbescheunigung, l = Pendellänge). Zur Zeit $t = 0$ soll $\varphi = 0$ sein und die Anfangsgeschwindigkeit $l\dot{\varphi}(0)$ betrage $v_0 > 0$. Gemäß der allgemeinen Theorie genügt $\varphi(t)$ der Dgl. 1. Ordnung $\dot{\varphi} = \pm[(2g/l)\cos\varphi + 2C]^{1/2}$. Aus den Anfangsbedingungen folgt, daß hier nur das obere Vorzeichen brauchbar ist und daß sich $2C = (v_0/l)^2 - (2g/l)$ ergibt. Also ist $\mathrm{d}\varphi/\mathrm{d}t = [(2g/l)(-1 + \cos\varphi) + (v_0/l)^2]^{1/2}$. Trennen der Veränderlichen führt zu $t = t(\varphi) = l\int_0^{\varphi} [2gl(-1 + \cos\tilde{\varphi}) + v_0^2]^{-1/2}\mathrm{d}\tilde{\varphi} + C_1$ mit $C_1 = 0$ wegen $\varphi(0) = 0$ und damit $t(\varphi) = 0$ für $\varphi = 0$. Das erhaltene Integral gehört zu den *elliptischen Integralen*, die i. allg. nicht durch elementare Funktionen angebbar sind. Hier wird die gesuchte Funktion in der nach t aufgelösten Gestalt $t = t(\varphi)$ angegeben. Im Fall $v_0^2 > 4gl$ ist der Integrand stets positiv, also $t = t(\varphi)$ und damit auch $\varphi = \varphi(t)$ für alle t monoton wachsend. Das bedeutet ein ständiges Überschlagen des Pendels. Ist $v_0^2 < 4gl$, so ist die kleinste positive Nullstelle des Nenners im Integranden gleich $\varphi_1 = \arccos[1 - (v_0^2/2gl)]$. Das Pendel schwingt von $\varphi = 0$ bis zum Maximalausschlag φ_1, hat dort die Geschwindigkeit 0 und schwingt danach wieder zurück. Die benötigte Zeit T_1 für eine Viertelschwingung von $\varphi = 0$ bis $\varphi = \varphi_1$ beträgt $T_1 = l\int_0^{\varphi_1} [2gl(-1 + \cos\varphi) + v_0^2]^{-1/2}\mathrm{d}\varphi$. Bei „kleiner" Anfangsgeschwindigkeit v_0 kann man mit guter Näherung die Kosinusfunktion durch die ersten beiden von null verschiedenen Glieder ihrer Taylor-

Entwicklung (Entwicklungsstelle null) ersetzen. Damit ergibt sich für φ_1 näherungsweise $v_0(gl)^{-1/2}$ und für die Schwingungsdauer $T = 4T_1$ die bekannte Näherung $T = 2\pi(l/g)^{1/2}$.

13.5.3.3. Die Differentialgleichung $y'' = f(x, y')$. Die Dgl. $y'' = f(x, y')$ ist für $p(x) = y'(x)$ die Dgl. 1. Ordnung $p' = f(x, p)$. Bei bekanntem $p(x)$ ergibt sich $y(x)$ durch unbestimmte Integration.

13.5.3.4. Die Differentialgleichung $y'' = f(y, y')$. Im Fall $y'(x) \neq 0$ denkt man sich die Lösung $y = y(x)$ der Dgl. $y'' = f(y, y')$ nach x aufgelöst, $x = x(y)$, und danach $x = x(y)$ in $p(x) = y'(x)$ eingesetzt. Es entsteht $q(y) = p(x)$ mit $x = x(y)$. Damit ergibt sich für $q(y)$ die Dgl. 1. Ordnung $dq/dy = q^{-1}f(y, q)$. Bei bekanntem $q(y)$ erhält man $y(x)$ aus $y' = q(y)$ durch Trennen der Veränderlichen.

13.6. Gewöhnliche Differentialgleichungssysteme

Durch $F_\mu(x, y_1, y_1', \ldots, y_1^{(n_1)}, \ldots, y_m, y_m', \ldots, y_m^{(n_m)}) = 0$ $(\mu = 1, \ldots, m)$ ist ein *gewöhnliches Differentialgleichungssystem (gekoppelte Differentialgleichungen, simultane Differentialgleichungen)* für ein Funktionen-m-tupel $[y_1(x), \ldots, y_m(x)]$ gegeben. Kommen die Ableitungen höchster Ordnung $y_1^{(n_1)}, \ldots, y_m^{(n_m)}$ in mindestens einer der Gleichungen vor, so heißt das Differentialgleichungssystem bezüglich y_1 von n_1-ter Ordnung, ..., bezüglich y_m von n_m-ter Ordnung. Die Summe $n_1 + \ldots + n_m$ gibt die *Ordnung des Differentialgleichungssystems* an.

13.6.1. Gewöhnliche lineare Differentialgleichungssysteme

Man erhält aus $\sum_{\nu=0}^{n} a_\nu(x) y^{(\nu)} = g(x)$ (vgl. 13.5.1.) ein explizites gewöhnliches lineares Differentialgleichungssystem von m Gleichungen für das Funktionen-m-tupel $[y_1(x), y_2(x), \ldots, y_m(x)]$, wobei das Differentialgleichungssystem bezüglich y_μ $(\mu = 1, \ldots, m)$ jeweils von n-ter Ordnung ist, insgesamt also ein Differentialgleichungssystem von $n \cdot m$-ter Ordnung vorliegt, auf folgende Weise: Man ersetzt $y^{(\nu)}(x)$ $(\nu = 1, \ldots, n)$ und $g(x)$ durch Matrizen $\boldsymbol{y}^{(\nu)}(x)$ und $\boldsymbol{g}(x)$ vom Typ $(m, 1)$, wobei $\boldsymbol{y}^{(\nu)}(x)$ die Elemente $y_\mu^{(\nu)}(x)$ $(\mu = 1, \ldots, m)$ und $\boldsymbol{g}(x)$ die Elemente $g_\mu(x)$ $(\mu = 1, \ldots, m)$ aufweisen. Die Koeffizienten $a_\nu(x)$ werden durch Matrizen $\boldsymbol{A}_\nu(x)$ vom Typ (m, m) ersetzt. Die hierdurch entstehenden Produkte zwischen $\boldsymbol{A}_\nu(x)$ und $\boldsymbol{y}^{(\nu)}(x)$ sind im Sinne der Matrizenmultiplikation aufzufassen. Die Forderung $a_n(x) \neq 0$ wird hier durch $\det \boldsymbol{A}_n(x) \neq 0$ ersetzt. Sind alle Elemente des *Störgliedes* $\boldsymbol{g}(x)$ gleich null, so spricht man von einem *homogenen* linearen Differentialgleichungssystem, andernfalls von einem *inhomogenen* Differentialgleichungssystem.

13.6.1.1. Lösungsstruktur linearer Differentialgleichungssysteme. Analog zu 13.5.1.1. gilt: Sämtliche Lösungen werden in der *allgemeinen Lösung* angegeben. Diese ist gleich einer *partikulären* Lösung des inhomogenen Differentialgleichungssystems plus der *allgemeinen* Lösung des zugehörigen homogenen Differentialgleichungssystems. Sämtliche Lösungen des zugehörigen homogenen Differentialgleichungssystems bilden einen Vektorraum, dessen Dimension mit der Ordnung des Differentialgleichungssystems übereinstimmt.

13.6.1.2. Lineare homogene Differentialgleichungssysteme mit konstanten Koeffizienten. Um eine Basis des $n \cdot m$-dimensionalen Lösungsraumes des homogenen Differentialgleichungssystems $\sum_{\nu=0}^{n} \boldsymbol{A}_\nu \boldsymbol{y}_h^{(\nu)} = 0$ ($\boldsymbol{A}_\nu$ hat konstante Elemente, $\det \boldsymbol{A}_n \neq 0$) zu erhalten, setzt man zunächst den Ansatz $\boldsymbol{y}_h = \boldsymbol{d}\mathrm{e}^{\lambda x}$ $(\boldsymbol{d} \neq 0)$ ein, wobei $\boldsymbol{d}$ eine konstante unbekannte Matrix vom Typ $(m, 1)$ ist. Dividiert man danach die Gleichung durch $\mathrm{e}^{\lambda x}$, so ergibt sich für die Elemente von $\boldsymbol{d}$ das lineare homogene Gleichungssystem (nicht: Differentialgleichungssystem) $\left(\sum_{\nu=0}^{n} \boldsymbol{A}_\nu \lambda^\nu\right)\boldsymbol{d} = \boldsymbol{0}$. Lösungen $\boldsymbol{d} \neq \boldsymbol{0}$ liegen nur vor, wenn die Koeffizientendeterminante gleich null ist, d. h. wenn λ der *charakteristischen Gleichung* $\det\left(\sum_{\nu=0}^{n} \boldsymbol{A}_\nu \lambda^\nu\right) = 0$ genügt. Die Lösung λ_k der charakteristischen Gleichung mit der Vielfachheit l_k $\left(\sum_k l_k = m \cdot n\right)$ wird in das Gleichungssystem für die Elemente von $\boldsymbol{d}$ eingesetzt: $\left(\sum_{\nu=0}^{n} \boldsymbol{A}_\nu \lambda_k^\nu\right)\boldsymbol{d} = \boldsymbol{0}$. Wenn der Lösungsraum dieses Gleichungssystems für $\boldsymbol{d}$ die Dimension l_k hat mit den Basiselementen $\boldsymbol{d}_{k\varrho}$ $(\varrho = 1, \ldots, l_k)$, so sind damit l_k Basiselemente $\boldsymbol{d}_{k\varrho}\exp(\lambda_k x)$ $(\varrho = 1, \ldots, l_k)$ des Lösungsraumes vom Differentialgleichungssystem gefunden.

Sollte jedoch die Dimension des $\boldsymbol{d}$-Lösungsraumes gleich $r_k < l_k$ sein, so erhält man zunächst nur r_k Basiselemente des Lösungsraumes des Differentialgleichungssystems. Die noch fehlenden restlichen $l_k - r_k$ Basiselemente haben die kompliziertere Struktur $(\boldsymbol{d}^{(0)} + \boldsymbol{d}^{(1)}x + \ldots + \boldsymbol{d}^{l_k - 1}x^{l_k - 1})\mathrm{e}^{\lambda_k x}$. Diese Formel wird in das Differentialgleichungssystem eingesetzt, danach die entstehende Matrizengleichung durch $\mathrm{e}^{\lambda_k x}$ dividiert und schließlich ein Koeffizientenvergleich durchgeführt. Es entsteht ein lineares homogenes Gleichungssystem von $l_k \cdot m$ Gleichungen für die $l_k \cdot m$ Elemente von $\boldsymbol{d}^{(0)}, \ldots, \boldsymbol{d}^{(l_k - 1)}$. Es zeigt sich, daß der Lösungsraum dieses Gleichungssystems die Dimension l_k hat. Man ist daher in der Lage, die noch fehlenden $l_k - r_k$ Basiselemente des Differentialgleichungssystems in der Gestalt $(\boldsymbol{d}_{k_\varrho}^{(0)} + \boldsymbol{d}_{k_\varrho}^{(1)}x + \ldots + \boldsymbol{d}_{k_\varrho}^{l_k - 1}x^{l_k - 1})\mathrm{e}^{\lambda_k x}$ $(\varrho = r_k + 1, \ldots, l_k)$ anzugeben. Sind alle Koeffizienten des Differentialgleichungssystems reell, so ist mit jeder nicht-reellen Lösung λ_k der charakteristischen Gleichung auch λ_k^* eine Lösung. Insbesondere kann man aus den zu λ_k gehörigen Basiselementen des Lösungsraumes des Differentialgleichungssystems durch Übergang zum Konjugiert-Komplexen Basiselemente erhalten, die zu λ_k^* gehören. Bildet man dann analog zu 13.5.1.2. von zueinander konjugiert-komplexen Basiselementen einerseits die „halbe Summe" und andererseits die mit 1/i zu multiplizierende „halbe Differenz", so ist der Übergang zur *reellen Basis* vollzogen.

Bsp. 13.6.1–1. Zwei Punktmassen m_1 und m_2 befinden sich an den Stellen 0_1 und 0_2 $(0_1 < 0_2)$ einer Zahlengeraden in Ruhe und sind durch eine entspannte Feder (Federkonstante $c > 0$) verbunden. Zum Zeitpunkt $t = 0$ erfährt m_1 eine Anfangsgeschwindigkeit v_0 in Richtung der Zahlengeraden. Gesucht sind die Bewegungen $x_1 = x_1(t)$ und $x_2 = x_2(t)$, wobei x_1 bzw. x_2 den orientierten Abstand der Massen m_1 bzw. m_2 vom Punkt 0_1 bzw. 0_2 messen. Es ist somit die folgende Anfangswertaufgabe zu lösen:

$m_1\ddot{x}_1 + c(x_1 - x_2) = 0$, $m_2\ddot{x}_2 + c(x_2 - x_1) = 0$, $x_1(0) = x_2(0) = \dot{x}_2(0) = 0$, $\dot{x}_1(0) = v_0$ $(v_0 > 0)$. Dieses Differentialgleichungssystem lautet in Matrizenschreibweise $\boldsymbol{A}\ddot{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{x} = \boldsymbol{0}$

mit

$$A = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix}, \quad B = \begin{pmatrix} c & -c \\ -c & c \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad 0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Der Ansatz $x = d e^{\lambda t}$ führt zu $(A\lambda^2 + B)d = 0$. Die Lösungen der charakteristischen Gleichung $\det(A\lambda^2 + B) = 0$ lauten (mit der Abkürzung $\omega = [c(m_1 + m_2)]^{1/2}(m_1 m_2)^{-1/2}$, $\lambda_1 = 0$ (Vielfachheit $l_1 = 2$), $\lambda_2 = i\omega$ $(l_2 = 1)$, $\lambda_3 = -i\omega$ $(l_3 = 1)$. Wegen $l_1 = 2$ muß gemäß der allgemeinen Theorie $\lambda_1 = 0$ *zwei* Basiselemente des Lösungsraumes des Differentialgleichungssystems liefern. Da jedoch das zu $\lambda_1 = 0$ gehörige Gleichungssystem für die Elemente von d, nämlich $Bd = 0$, nur einen *ein*dimensionalen Lösungsraum aufweist, ist gemäß der allgemeinen Theorie der erweiterte Ansatz $x(t) = d^{(0)} + d^{(1)}t$ erforderlich. Einsetzen in das Differentialgleichungssystem und Koeffizientenvergleich liefert das – hier speziell in zwei Gleichungssysteme für $d^{(0)}$ und $d^{(1)}$ zerfallende – Gleichungssystem $Bd^{(0)} = 0$, $Bd^{(1)} = 0$. Die Dimension des zu diesem Gleichungssystem von vier Gleichungen für die vier Elemente von $d^{(0)}$ und $d^{(1)}$ gehörenden Lösungsraumes ist gleich der Vielfachheit l_1 von λ_1 und damit gleich zwei. Als Basis können die beiden Matrizen vom Typ (4,1) angegeben werden: $(d_{11}^{(0)\mathrm{T}} d_{11}^{(1)\mathrm{T}})^{\mathrm{T}} = (1\,1\,0\,0)^{\mathrm{T}}$ und $(d_{12}^{(0)\mathrm{T}} d_{12}^{(1)\mathrm{T}})^{\mathrm{T}} = (0\,0\,1\,1)^{\mathrm{T}}$. Damit liefert schließlich $\lambda_1 = 0$ für den Lösungsraum des Differentialgleichungssystems die beiden Basiselemente $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ und $\begin{pmatrix} 1 \\ 1 \end{pmatrix} t$. Für $\lambda_2 = i\omega$ lautet das d-Gleichungssystem $(-A\omega^2 + B)d = 0$. Als Basis des zugehörigen eindimensionalen Lösungsraumes kann man $\begin{pmatrix} m_2 \\ -m_1 \end{pmatrix}$ nehmen. Damit wird als weiteres Basiselement des Lösungsraumes des Differentialgleichungssystems $\begin{pmatrix} m_2 \\ -m_1 \end{pmatrix} e^{i\omega t}$ geliefert. Zu $\lambda_3 = -i\omega$ ergibt sich analog das Basiselement $\begin{pmatrix} m_2 \\ -m_1 \end{pmatrix} e^{-i\omega t}$. Geht man danach zu einer reellen Basis über, so erhält man schließlich durch Linearkombination der berechneten vier reellen Basiselemente die allgemeine Lösung des Differentialgleichungssystems:

$$x(t) = C_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + C_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} t + C_3 \begin{pmatrix} m_2 \\ -m_1 \end{pmatrix} \cos(\omega t)$$

$$+ C_4 \begin{pmatrix} m_2 \\ -m_1 \end{pmatrix} \sin(\omega t).$$ Mittels der Anfangsbedingungen werden $C_1, \ldots, C_4$ bestimmt. Damit lautet die Lösung der gegebenen Anfangswertaufgabe:

$x_1(t) = v_0(m_1 + m_2)^{-1}[m_1 t + m_2(\omega)^{-1}\sin(\omega t)]$,

$x_2(t) = v_0(m_1 + m_2)^{-1}[m_1 t - m_1(\omega)^{-1}\sin(\omega t)]$.

13.6.1.3. Partikuläre Lösung des linearen inhomogenen Differentialgleichungssystems mit konstanten Koeffizienten. Ansatzmethode. Hat das Störglied $g(x)$ des Differentialgleichungssystems $\sum_{\nu=0}^{n} A_\nu y^{(\nu)} = g(x)$ (A_ν hat konstante Elemente; $\det A_n \neq 0$) die Struktur $g(x) = \sum_{\sigma=0}^{s} b_\sigma x^\sigma e^{qx}$ (b = Matrix vom Typ $(m, 1)$, $b_s \neq 0$), so macht man für eine partikuläre Lösung $y_p(x)$ den – im Vergleich zu 13.5.1.3. komplizierteren – Ansatz $y_p(x) = \sum_{\mu=0}^{\sigma+l} B_\mu x^\mu e^{qx}$ mit $l = 0$, falls q keine Lösung der charakteristischen Gleichung $\det \sum_{\nu=0}^{n} A_\nu \lambda^\nu = 0$ ist; andernfalls soll l gleich der Vielfachheit der Lösung q der charakteristischen Gleichung sein. Zur Bestimmung von $B_0, B_1, \ldots, B_{\sigma+l}$ setzt man den Ansatz ins Differentialgleichungssystem ein, dividiert anschließend beide Seiten durch e^{qx} und führt schließlich einen Koeffizientenvergleich durch. Es ergibt sich ein lineares Gleichungssystem von $m(\sigma + l + 1)$ Gleichungen für die $m(\sigma + l + 1)$ Elemente von $B_0, \ldots, B_{\sigma+l}$, das genau eine Lösung hat.

Erweiterung der Ansatzmethode und Superponieren. Der Text aus 13.5.1.3. ist hierzu unmittelbar übertragbar.

Bsp. 13.6.1–2. Zur Bestimmung einer partikulären Lösung des Differentialgleichungssystems $y_1' - 4y_1 - y_2 = 3x$, $y_2' + 2y_1 - y_2 = e^{-x}$ schreibt man es zunächst in der Matrizengestalt $Ay' + By = g(x)$, ermittelt aus der charakteristischen Gleichung $\det(A\lambda + B) = 0$ die Lösungen $\lambda_1 = 2$, $\lambda_2 = 3$, zerlegt das Störglied $g(x)$ gemäß $g(x) = g_{\mathrm{I}}(x) + g_{\mathrm{II}}(x)$ mit $g_{\mathrm{I}}(x) = \begin{pmatrix} 3 \\ 0 \end{pmatrix} x$ $g_{\mathrm{II}}(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} e^{-x}$ und erhält die partikuläre Lösung $y_p(x) = y_{\mathrm{I}}(x) + y_{\mathrm{II}}(x)$, wobei $Ay_{\mathrm{I}}' + By_{\mathrm{I}} = g_{\mathrm{I}}(x)$, $Ay_{\mathrm{II}}' + By_{\mathrm{II}} = g_{\mathrm{II}}(x)$ gilt. Da $q = 0$ und $q = -1$ keine Lösungen der charakteristischen Gleichung sind, werden die Ansätze $y_{\mathrm{I}}(x) = B_{0\mathrm{I}} + B_{1\mathrm{I}}x$ und $y_{\mathrm{II}}(x) = B_{0\mathrm{II}}e^{-x}$ gemacht. Für $B_{0\mathrm{I}}, B_{1\mathrm{I}}, B_{0\mathrm{II}}$ ergeben sich die Gleichungssysteme $AB_{1\mathrm{I}} + BB_{0\mathrm{I}} = 0$, $BB_{0\mathrm{I}} = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$, $-AB_{0\mathrm{II}} + BB_{0\mathrm{II}} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Hieraus folgt $B_{0\mathrm{I}} = \begin{pmatrix} 1/12 \\ -5/6 \end{pmatrix}$, $B_{1\mathrm{I}} = \begin{pmatrix} -1/2 \\ -1 \end{pmatrix}$, $B_{0\mathrm{II}} = \begin{pmatrix} 1/12 \\ -5/12 \end{pmatrix}$.

Also lautet die gesuchte partikuläre Lösung

$y_1(x) = (1/12) - (1/2)x + (1/12)\exp(-x)$,

$y_2(x) = (-5/6) - x - (5/12)\exp(-x)$.

Variation der Konstanten (auch bei *variablen* Koeffizienten gültig). Eine partikuläre Lösung $y_p(x)$ von $\sum_{\nu=0}^{n} A_\nu y^{(\nu)} = g(x)$ $(\det A_n \neq 0)$ kann man mit dem Ansatz $y_p(x) = \sum_{\mu=1}^{m \cdot n} u_\mu(x) y_{h\mu}(x)$ finden, wobei die Matrizen $y_{h\mu}(x)$ vom Typ $(m, 1)$ $(\mu = 1, \ldots, m \cdot n)$ eine Basis des $m \cdot n$-dimensionalen Lösungsraumes des zugehörigen homogenen Differentialgleichungssystems $\sum_{\nu=0}^{n} A_\nu y_h^{(\nu)} = 0$ bilden. Die Ableitungen $u_1'(x), \ldots, u_{m \cdot n}'(x)$ ergeben sich als Lösungen des folgenden aus $m \cdot n$ Gleichungen bestehenden linearen inhomogenen Gleichungssystems (nicht Differentialgleichungssystem) $\sum_{\mu=1}^{m \cdot n} u_\mu'(x) y_{h\mu}^{(\nu)}(x) = 0$ $(\nu = 0, \ldots, n-2)$, $\sum_{\mu=1}^{m \cdot n} u_\mu'(x) y_{h\mu}^{(n-1)}(x) = A^{-1} g(x)$. Da nur ein y_p benötigt wird, kann bei der Bestimmung von $u_\mu(x)$ aus $u_\mu'(x)$ auf die Integrationskonstante verzichtet werden.

13.6.2. Zurückführen auf Differentialgleichungssysteme, die nur Ableitungen erster Ordnung enthalten

Führt man im Differentialgleichungssystem aus 13.6., das aus m Dgl.en besteht und die Ordnung $n_1 + \ldots + n_m$ hat, für die $n_1 + \ldots + n_m$ Funktionen $y_1, y_1', \ldots, y_1^{(n_1-1)}, \ldots, y_m, y_m', \ldots, y_m^{(n_m-1)}$ der Reihe nach die Bezeichnungen $w_1(x), w_2(x), \ldots, w_{n_1}(x), \ldots, w_{n_1+\ldots+n_m}(x)$ ein, so ergibt sich für das $(n_1 + \ldots + n_m)$-tupel $(w_1, \ldots, w_{n_1+\ldots+n_m})$ das folgende dem gegebenen Differentialgleichungssystem $F_\mu(x, y_1, y_1', \ldots, y_1^{(n_1)}, \ldots, y_m, y_m', \ldots, y_m^{(n_m)}) = 0$ $(\mu = 1, \ldots, m)$ der Ordnung $n_1 + \ldots + n_m$ äquivalente Differentialgleichungssy-

stem der Ordnung $n_1 + \ldots + n_m$, das aus $n_1 + \ldots + n_m$ Dgl.en besteht und nur Ableitungen erster Ordnung enthält: $F_\mu(x, w_1, w_2, \ldots, w_{n_1}, w'_{n_1}, w_{n_1+1}, \ldots, w_{n_1+n_2}, w'_{n_1+n_2}, \ldots, w_{n_1+\ldots+n_m}, w'_{n_1+\ldots+n_m}) = 0$, $(\mu = 1, \ldots, m)$,
$w'_1 - w_2 = 0, w'_2 - w_3 = 0, \ldots, w'_{n_1-1} - w_{n_1} = 0$,
$w'_{n_1+1} - w_{n_1+2} = 0, \ldots, w'_{n_1+n_2-1} - w_{n_1+n_2} = 0, \ldots$,
$w'_{n_1+\ldots+n_m-1} - w_{n_1+\ldots+n_m} = 0$.

13.6.3. Zurückführen eines Differentialgleichungssystems auf eine Differentialgleichung höherer Ordnung

Ein gegebenes Differentialgleichungssystem für $(y_1, \ldots, y_m)$, das nur Ableitungen erster Ordnung enthält, liege in der expliziten d. h. nach $y'_1, \ldots, y'_m$ aufgelösten Gestalt vor: $y'_\mu = f_\mu(x, y_1, \ldots, y_m)$ $(\mu = 1, \ldots, m)$. Man denkt sich eine Lösung $[y_1(x), \ldots, y_m(x)]$ in die erste Dgl. dieses Differentialgleichungssystems eingesetzt und danach die entstandene Identität $m-1$ mal nach x differenziert (Einsteinsche Summationsverabredung mit $k = 1, \ldots, f$):

$$y''_1 = \frac{\partial f_1}{\partial x} + \sum_{k=1}^{m} \frac{\partial f_1}{\partial y_k} y'_k = \frac{\partial f_1}{\partial x} + \sum_{k=1}^{m} \frac{\partial f_1}{\partial y_k} f_k$$

kurz: gleich $g_2(x, y_1, \ldots, y_m)$,

$$y'''_1 = \frac{\partial g_2}{\partial x} + \sum_{k=1}^{m} \frac{\partial g_2}{\partial y_k} f_k \text{ kurz: gleich } g_3(x, y_1, \ldots, y_m), \ldots,$$

$$y_1^{(m)} = \frac{\partial g_{m-1}}{\partial x} + \sum_{k=1}^{m} \frac{\partial g_{m-1}}{\partial y_k} f_k \text{ kurz: gleich } g_m(x, y_1, \ldots, y_m).$$

Von den Ergebnissen greift man die folgenden $m-1$ Gleichungen heraus: $y'_1 = f_1(x, y_1, \ldots, y_m)$, $y_1^{(\mu)} = g_\mu(x, y_1, \ldots, y_m)$ $(\mu = 2, \ldots, m-1)$. Man *setzt voraus*, daß dieses System (hier nicht Differentialgleichungssystem) nach den $m-1$ Größen $y_2, \ldots, y_m$ eindeutig auflösbar sei: $y_\mu = \varphi_\mu(x, y_1, y'_1, \ldots, y_1^{(m-1)})$ $(\mu = 2, \ldots, m)$. Diese $m-1$ Gleichungen setzt man in $y_1^{(m)} = g_m(x, y_1, \ldots, y_m)$ ein und erhält für $y_1(x)$ die Dgl. m-ter Ordnung
$y_1^{(m)} = g_m(x, y_1, \varphi_2(x, y_1, \ldots, y_1^{(m-1)}), \ldots, \varphi_m(x, y_1, \ldots, y_1^{(m-1)}))$,
kurz $y_1^{(m)} = \Phi(x, y_1, y'_1, \ldots, y_1^{(m-1)})$. Hat man aus dieser Dgl. $y_1(x)$ bestimmt, so ergeben sich die übrigen Funktionen $y_2(x), \ldots, y_m(x)$ durch Einsetzen von $y_1(x)$ in $y_\mu = \varphi_\mu(x, y_1, y'_1, \ldots, y_1^{(m-1)})$.

13.6.4. Kanonische Differentialgleichungssysteme

Mit den verallgemeinerten Koordinaten q_k und den verallgemeinerten Impulsen p_k $(k = 1, \ldots, f$ bei f Freiheitsgraden) – je zwei Größen q_k und p_k sind gepaart, sie sind *kanonisch-konjugiert* – wird für die Funktionen $q_k(t)$, $p_k(t)$ (Zeit t) das Differentialgleichungssystem

$$\dot{q}_k = \frac{\partial H}{\partial p_k}(t, q_1, \ldots, q_f, p_1, \ldots, p_f),$$

$$\dot{p}_k = -\frac{\partial H}{\partial q_k}(t, q_1, \ldots, q_f, p_1, \ldots, p_f)$$

mit der *Hamilton-Funktion* $H(t, q_1, \ldots, q_f, p_1, \ldots, p_f)$ – kurz $H(t, q_l, p_l)$ – gebildet. Setzt man eine Lösung $[q_1(t), \ldots, q_f(t), p_1(f), \ldots, p_f(t)]$ des Differentialgleichungssystems in $H(t, q_l, p_l)$ ein, so ergibt sich

$$\frac{d}{dt} H(t, q_l(t), p_l(t)) = \frac{\partial H}{\partial t}(t, q_l, p_l) + \sum_{l=1}^{f} \left(\frac{\partial H}{\partial q_l} \dot{q}_l + \frac{\partial H}{\partial p_l} \dot{p}_l \right),$$

wobei die letzte Summe wegen $\dot{q}_l = \frac{\partial H}{\partial p_l}, \dot{p}_l = -\frac{\partial H}{\partial q_l}$ stets gleich null ist. Ist insbesondere H von t explizit nicht abhängig, d. h. $H = H(q_l, p_l)$, so ist $dH/dt = 0$ und damit $H(q_l(t), p_l(t)) =$ konst. (Energie ist Erhaltungsgröße).

13.6.4.1. Kanonische Transformation. Es handelt sich um Transformationen $Q_k = Q_k(t, q_1, \ldots, q_f, p_1, \ldots, p_f)$, $P_k = P_k(t, q_1, \ldots, q_f, p_1, \ldots, p_f)$ – kurz: $Q_k = Q_k(t, q_l, p_l)$, $P_k = P_k(t, q_l, p_l)$ – die eindeutig nach q_k, p_k $(k = 1, \ldots, f)$ auflösbar sind und das kanonische Differentialgleichungssystem $\dot{q}_k = \partial H/\partial p_k$, $\dot{p}_k = -\partial H/\partial q_k$ wieder in ein kanonisches Differentialgleichungssystem $\dot{Q}_k = \partial\tilde{H}/\partial P_k$, $\dot{P}_k = -\partial\tilde{H}/\partial Q_k$ mit der *neuen* Hamilton-Funktion $\tilde{H} = \tilde{H}(t, Q_l, P_l)$ überführen.

13.6.4.2. Zum kanonischen Differentialgleichungssystem gehörige Variationsaufgabe. Die Eulerschen Dgl.en an der Variationsaufgabe (Summationsverabredung $k = 1, \ldots, f$)

$$\delta \int_{t_1}^{t_2} L(t, q_1, \ldots, q_f, p_1, \ldots, p_f, \dot{q}_1, \ldots, \dot{q}_f, \dot{p}_1, \ldots, \dot{p}_f) dt = 0$$

(vgl. 13.3.3.10.) mit $L = p_k \dot{q}_k - H(t, q_1, \ldots, q_f, p_1, \ldots, p_f)$ ergeben das kanonische Differentialgleichungssystem $\dot{q}_k = \partial H/\partial p_k$, $\dot{p}_k = -\partial H/\partial q_k$. Analog hierzu gehört zum transformierten kanonischen Differentialgleichungssystem aus 13.6.4.1. die Variationsaufgabe

$$\delta \int_{t_1}^{t_2} \tilde{L}(t, Q_1, \ldots, Q_f, P_1, \ldots, P_f, \dot{Q}_1, \ldots, \dot{Q}_f, \dot{P}_1, \ldots, \dot{P}_f) dt = 0 \text{ mit}$$

$$\tilde{L} = P_k \dot{Q}_k - \tilde{H}(t, Q_1, \ldots, Q_f, \ldots, P_1, \ldots, P_f).$$

13.6.4.3. Erzeugende Funktion kanonischer Transformationen. Schreibt man bei den Variationsaufgaben aus 13.6.4.2. vor, daß $q_k(t)$, $p_k(t)$ bzw. $Q_k(t)$, $P_k(t)$ an den Integrationsgrenzen $t = t_1$ und $t = t_2$ jeweils vorgegebene Werte annehmen, so werden einerseits die Eulerschen Differentialgleichungssysteme der Variationsaufgaben, also hier die kanonischen Differentialgleichungssysteme nicht geändert und andererseits ist

$$\delta \int_{t_1}^{t_2} \frac{d}{dt} \{Q_k(t) P_k(t) + V_1[t, q_l(t), P_l(t)]\} dt$$

stets gleich null, welche Vergleichsfunktionen (vgl. 13.3.1.5.) $q_k(t)$, $Q_k(t)$, $P_k(t)$ und welche Funktion V_1 auch genommen werden, da sich der Wert des Integrals

$$\int_{t_1}^{t_2} \frac{d}{dt} \{Q_k(t) P_k(t) + V_1[t, q_l(t), P_l(t)]\} dt = Q_k(t_2) P_k(t_2)$$

$+ V_1[t_2, q_l(t_2), P_l(t_2)] - Q_k(t_1) P_k(t_1) - V_1[t_1, q_l(t_1), P_l(t_1)]$ beim Übergang zu anderen Vergleichsfunktionen nicht ändert. Beachtet man, daß $Q_k(t)$, $P_k(t)$ und $q_k(t)$, $p_k(t)$ durch die kanonische Transformation zusammenhängen, so kann für den obigen Ausdruck $Q_k(t) P_k(t) + V_1[t, q_l(t), P_l(t)]$ auch $V[t, q_l(t), P_l(t)]$ gesetzt werden, wobei V beliebig ist, denn V_1 ist beliebig. Beide in 13.6.4.2. angegebenen Variationsprobleme liefern die gleichen Lösungen. Sie werden beim ersten Problem durch t, q_k, p_k beim zweiten durch t, Q_k, P_k beschrieben. Damit erhält man gewiß kanonische Transformationen, wenn man $Q_k = Q_k(t, q_l, p_l)$, $P_k = P_k(t, q_l, p_l)$, $\tilde{H}(t, Q_l, P_l)$ bei Berücksichtigung der obigen Ausführungen und bei Beachtung von $\frac{d}{dt} V[t, q_l(t), P_l(t)] = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial q_k} \dot{q}_k + \frac{\partial V}{\partial P_k} \dot{P}_k$ und $P_k \dot{Q}_k = \frac{d}{dt}(P_k Q_k) - \dot{P}_k Q_k$ derart wählt, daß mit den Bezeichnungen aus 13.6.4.2.
$L[t, q_1(t), \ldots, q_f(t), p_1(t), \ldots, p_f(t), \dot{q}_1(t) \ldots \dot{q}_f(t), \dot{p}_1(t), \ldots, \dot{p}_f(t)] = \tilde{L}[t, Q_1(t), \ldots, Q_f(t), P_1(t), \ldots, P_f(t), \dot{Q}_1(t) \ldots, \dot{Q}_f(t), \dot{P}_1(t), \ldots, \dot{P}_f(t)] + \frac{d}{dt} V[t, q_l(t), P_l(t)]$, d. h. (falls man $\sum_{k=1}^{f} P_k Q_k$ zu dem

beliebigen V hinzufügt und das Ergebnis wiederum mit V bezeichnet) $p_k\dot{q}_k - H(t, q_l, p_l) = -\dot{P}_k Q_k - \tilde{H}(t, Q_l, P_l) + \frac{\partial V}{\partial t} + \left(\frac{\partial V}{\partial q_k}\dot{q}_k + \frac{\partial V}{\partial P_k}\dot{P}_k\right)$ gilt. Diese Gleichung ist gewiß erfüllt, falls man die kanonischen Transformationen durch $p_k = \frac{\partial V}{\partial q_k}(t, q_l, P_l)$, $Q_k = \frac{\partial V}{\partial P_k}(t, q_l, P_l)$ angibt. Für die neue Hamilton-Funktion $\tilde{H}$ folgt dann $\tilde{H}(t, Q_l, P_l) = H(t, q_l, p_l) + \frac{\partial V}{\partial t}(t, q_l, P_l)$, wobei man sich q_l und p_l mittels der oben angegebenen Formeln, die die kanonische Transformation definieren, als Funktionen von Q_l und P_l berechnet denken möge. Die Funktion V heißt *Erzeugende der kanonischen Transformation*. Da V beliebig ist, sind somit unendlich viele kanonische Transformationen gefunden.

13.6.4.4. Hamilton-Jacobische Differentialgleichung. Im Fall $\tilde{H}(t, Q_l, P_l) \equiv 0$ lautet das transformierte kanonische Differentialgleichungssystem (vgl. 13.6.4.1.) $\dot{Q}_k = 0$, $\dot{P}_k = 0$, d. h. $Q_k(t)$, $P_k(t)$ sind jeweils konstante Funktionen, übernehmen somit die Rolle der willkürlichen Konstanten (Erhaltungsgrößen) in der Lösungsformel des Problems. Man bezeichnet in diesem Fall die Erzeugende $V(t, q_l, P_l)$ durch $S(t, q_l, P_l)$ (*Wirkungsfunktion*). Es geht jetzt $\tilde{H} = H + \frac{\partial V}{\partial t}$ in $H(t, q_l, p_l) + \frac{\partial S}{\partial t}(t, q_l, P_l) = 0$ über, wobei wegen $p_k = \frac{\partial V}{\partial q_k}$ (vgl. 13.6.4.3.) jetzt p_l durch $\frac{\partial S}{\partial q_l}$ zu ersetzen ist. Es entsteht somit die *Hamilton-Jacobische Dgl.* $H\left(t, q_l, \frac{\partial S}{\partial q_l}\right) + \frac{\partial S}{\partial t} = 0$; sie ist eine partielle Dgl. für die von den unabhängigen Veränderlichen t, q_l $(l = 1, \dots, f)$ abhängende Funktion S. Aus $S = S(t, q_l, P_l)$ ist abzulesen, daß ferner eine Abhängigkeit des S von willkürlichen Konstanten $P_1, \dots, P_f$ zu fordern ist, wobei durch die weitere Forderung $\det\left(\frac{\partial^2 S}{\partial q_k P_l}\right) \neq 0$ gesichert wird, daß durch $p_k = \frac{\partial S}{\partial q_k}$, $Q_k = \frac{\partial S}{\partial P_k}$ tatsächlich eine kanonische Transformation geliefert wird. Durch diese beiden Formeln werden die Lösungen des Problems $q_k = q_k(t)$, $p_k = p_k(t)$ (P_k = konst., Q_k = konst.) implizit dargestellt, falls $S(t, q_l, P_l)$ – man sagt ein *vollständiges Integral* der Hamilton-Jacobischen Dgl. – bekannt ist.

Separationsansatz. Ist speziell $H\left(t, q_l, \frac{\partial S}{\partial q_l}\right) + \frac{\partial S}{\partial t}$ als Summe $\sum_{k=0}^{f} a_k$ darstellbar, wobei a_0 nur von t und $\frac{\partial S}{\partial t}$ und a_k nur von q_k und $\frac{\partial S}{\partial q_k}$ $(k = 1, \dots, f)$ abhängen, so führt der Ansatz $S = S_0(t) + \sum_{k=1}^{f} S_k(q_k)$ zu *gewöhnlichen* Dgl. für $S_0(t)$ und $S_k(q_k)$.

Register

Die Stichwörter sind alphabetisch nach der Buchstabenfolge in ihnen bis zum Komma geordnet; nach dem Komma beginnt ein neues Alphabet. Die Umlaute „ä", „ö" und „ü" sind wie „a", „o" bzw. „u" eingeordnet.
Griechische Buchstaben sind nach der Aussprache alphabetisiert worden, vorangestellte Zahlen ebenfalls. Der Buchstabe „ß" wurde wie „ss" behandelt. Besteht ein Stichwort aus Adjektiv und Substantiv, so ist das Adjektiv in der Regel - durch Komma abgetrennt - nachgestellt worden. Haben mehrere aufeinanderfolgende Stichwörter gleiche Buchstabenfolge vor dem Komma, so wurden sie vom zweiten Stichwort ab durch einen Gedankenstrich ersetzt. Der Gedankenstrich steht auch für den durch einen senkrechten Strich abgetrennten vorderen Teil eines vorhergehenden Stichwortes.

A

B

E

F

G

H

I

J

K

L

N

Q

R

S

Kleine Enzyklopädie Struktur der Materie

Herausgegeben von Ch. Weißmantel, R. Lenk, W. Forker und D. Linke
760 Seiten mit 650 Abbildungen und 80 Tafeln
Format: 145 mm × 215 mm
Leinen · DDR 19,80 M · Ausland 26,50 DM
Bestellangaben:
ISBN 3-323-00139-7
577 219 4 – KE Struktur der Materie

Die unbelebte Natur ist Existenzbedingung allen Lebens, und unerschöpflich ist ihr Formenreichtum von der Mikrowelt der Atome bis zu den Gebilden des Universums. Anspruchsvoll und doch verständlich behandelt die Kleine Enzyklopädie umfassend den Aufbau der Materie, ihre grundlegenden Gesetzmäßigkeiten, Erscheinungsformen und Eigenschaften ebenso wie die heutigen und künftigen Möglichkeiten der technischen Nutzung – ein Nachschlagewerk von geradezu einmaliger Vollständigkeit für Studierende, Fachexperten und jeden naturwissenschaftlich Interessierten.

Zu beziehen durch den Buchhandel.
Der DM-Preis versteht sich ausschließlich Mehrwertsteuer.



VEB BIBLIOGRAPHISCHES INSTITUT LEIPZIG
Postschließfach 130, Leipzig, DDR-7010

BI-Taschenlexikon Energie

Herausgegeben von W. Conrad unter Mitarbeit von G. Ackermann, F. Häfner, E. Kraus, S. Lepenies, R. Nitsch, H.-D. Rentsch, E. Schesky, K. Schulte, G. Sörgel und K.-W. Walter
346 Seiten mit 200 Textabbildungen, 16 Fototafeln und 1 Karte auf dem Vorsatz
Format: 120 mm × 190 mm
Leinen · DDR 15,– M · Ausland 15,– DM
Bestellangaben:
576 961 4 – TL Energie

Mit keiner wissenschaftlich-technischen Errungenschaft wird der Mensch in seiner Umwelt so häufig konfrontiert wie mit der Anwendung von verschiedenen Energieformen zur Kraft-, Licht- und Wärmeerzeugung.
Das vorliegende Taschenlexikon enthält etwa 1500 Stichwörter aus den Bereichen der Energetik wie Nutzenergiegewinnung, Energieumwandlung, Energietransport, Energiespeicherung u. a. Dabei werden auch die entsprechenden Maschinen und Anlagen erläutert.

Zu beziehen durch den Buchhandel.
Der DM-Preis versteht sich ausschließlich Mehrwertsteuer.



VEB BIBLIOGRAPHISCHES INSTITUT LEIPZIG
Postschließfach 130, Leipzig, DDR-7010

BI-Lexikon Virologie

Herausgegeben von E. Geißler
408 Seiten mit 117 Abbildungen, 35 Tabellen und 16 Schwarzweißtafeln
Format: 145 mm × 215 mm
Leinen · DDR 32,– M · Ausland 32,– DM
Bestellangaben:
ISBN 3-323-00014-5
577 826 2 – Lex Virologie

In den letzten Jahren hat die Virologie als eine der modernen medizinisch-naturwissenschaftlichen Disziplinen eine stürmische Entwicklung und inhaltliche Erweiterung erfahren. Die Entdeckung vieler neuer, bisher unbekannter Viren und die Einbeziehung neuer Erkenntnisse und Methoden der Molekularbiologie und molekularen Genetik ließen die Virologie zu einem der attraktivsten Forschungsgebiete werden. Neben der praktischen Bedeutung für Human- und Veterinärmedizin sowie Phytopathologie haben Viren als wichtige Hilfsmittel des Genetic Engineering das Interesse breiter Kreise geweckt. Das Nachschlagewerk gibt naturwissenschaftlich interessierten Lesern die notwendige Hintergrundinformation, um die Bedeutung der neusten Ergebnisse aus dem Bereich der Virologie richtig einschätzen zu können.

Aus dem Inhalt:
AIDS, Biologische Waffen, Containment, Gensynthese, Gentechnik, Influenza, Herpesviren, Immunität, Immunreaktionen, infektiöses Aerosol, Interferone, Leukämieviren, Maul- und Klauenseuche, molekulare Vehikel, Parainfluenzaviren, Röteln-Virus, Tumorviren, Virosen, Virusinfektion

Zu beziehen durch den Buchhandel.
Der DM-Preis versteht sich ausschließlich Mehrwertsteuer.



VEB BIBLIOGRAPHISCHES INSTITUT LEIPZIG
Postschließfach 130, Leipzig, DDR-7010

BI-Lexikon Toxikologie

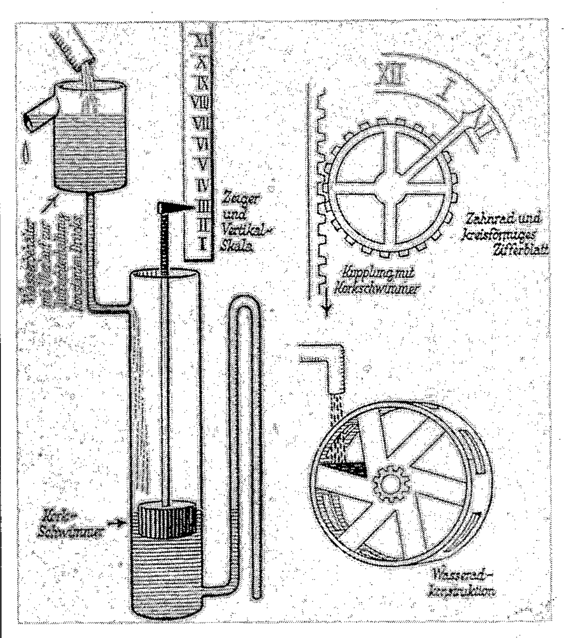
Herausgegeben von U. Stephan, P. Elstner und R. K. Müller
384 Seiten mit 161 teils zweifarbigen Abbildungen, 16 Farbtafeln und 34 Tabellen
Format: 145 mm × 215 mm
Leinen · DDR 32,– M · Ausland 32,– DM
Bestellangaben:
577 713 9 – Lexikon Toxikologie

Mit diesem Nachschlagewerk wird allen, die mit toxikologischen Fragen zu tun haben, eine Informationshilfe gegeben, in der die in der Literatur verstreuten Angaben über Gifte, ihre Charakteristika und die beim Umgang zu beachtenden Regeln in alphabetischer Folge behandelt werden. Der Benutzer erfährt etwas über die Wirksamkeit giftiger Pflanzen und Tiere, über Halluzinogene und Drogen, toxische Umweltchemikalien, Schadstoffausbreitung und Schadenverhütung, Industrietoxikologie und giftrechtliche Bestimmungen.

Zu beziehen durch den Buchhandel.
Der DM-Preis versteht sich ausschließlich Mehrwertsteuer.



VEB BIBLIOGRAPHISCHES INSTITUT LEIPZIG
Postschließfach 130, Leipzig, DDR-7010

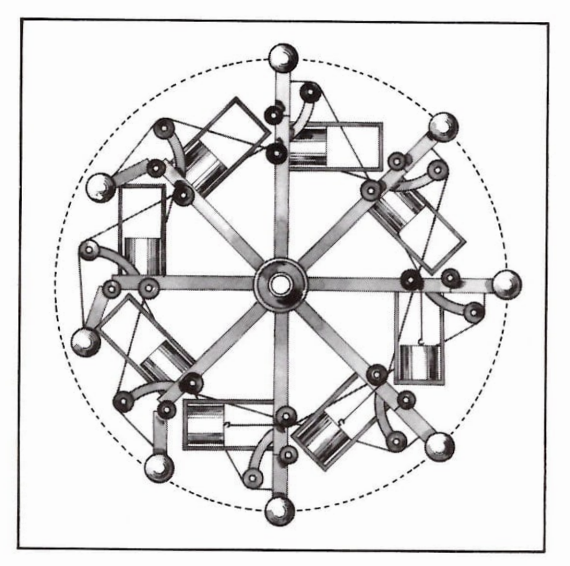


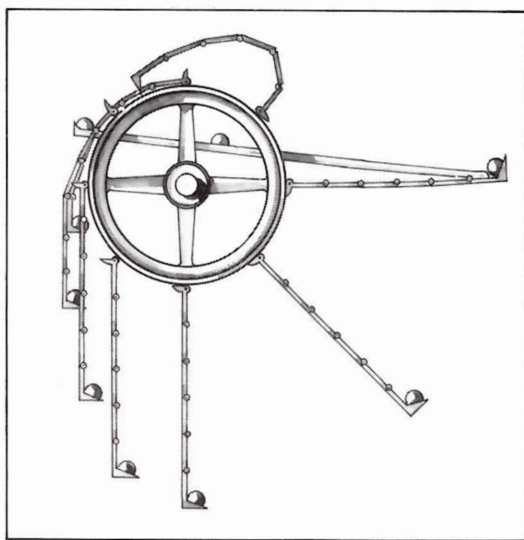


1 2

3

Tafel 1
1 Beispiele antiker griechischer Wasseruhren – 2 Demokrit von Abdera (460–370 v. u. Z.) – 3 Darstellung der Anwendung des Hebels beim Transport einer Statue im Assyrischen Reich.

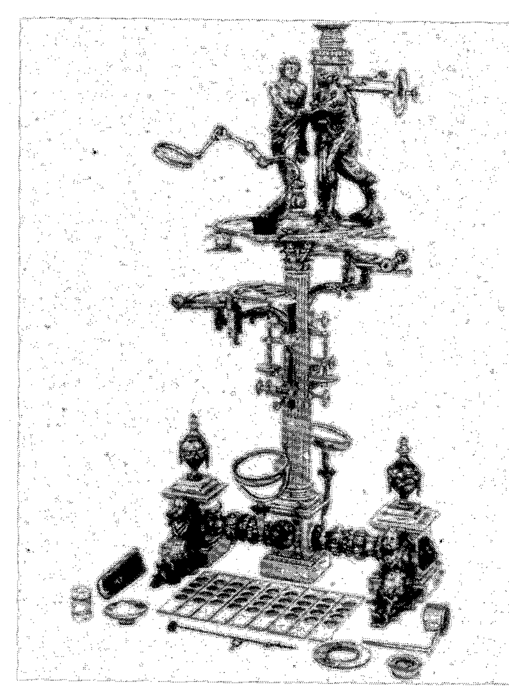


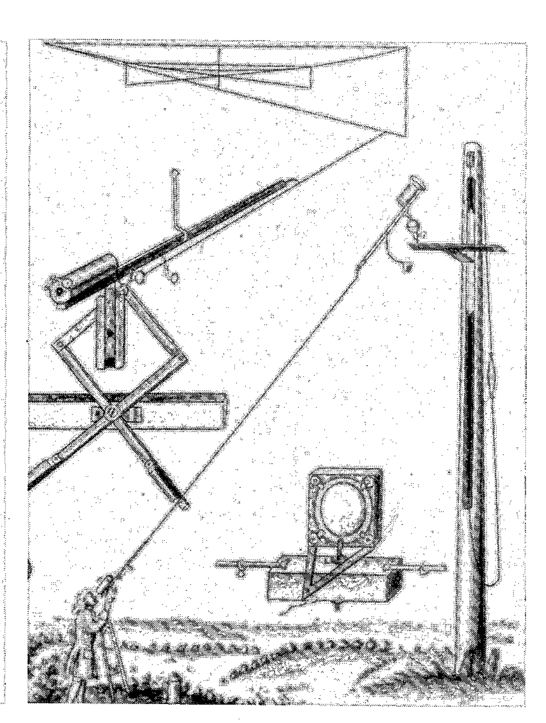


1 2 | 4 5
3 | 6

Tafeln 2 und 3
1–3 Beispiele für Perpetuum-Mobile-Konstruktionen – **4** Skizzen, die Galilei von verschiedenen Mondphasen anfertigte – **5.** Galileo Galilei (1564–1642) – **6** Von Galilei 1638 aufgestellte Wurfparabel.

E D C B
A
J O
F G
H L
N



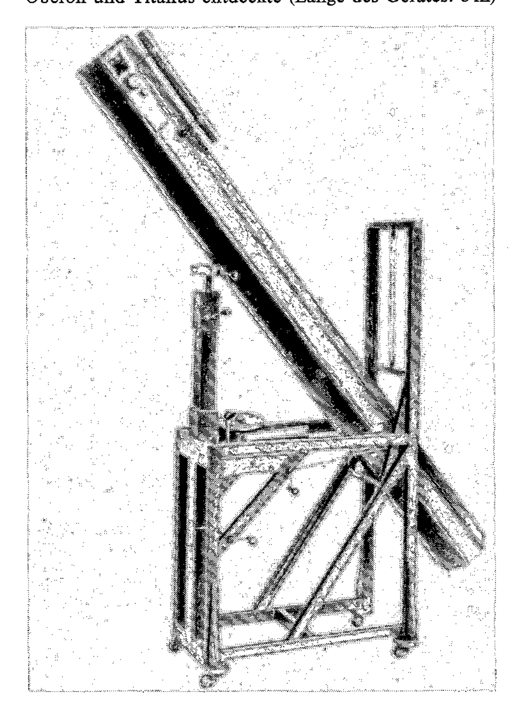


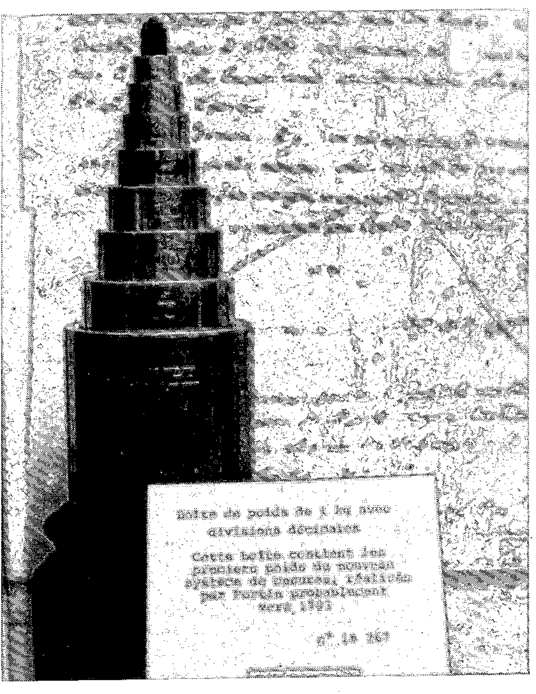
1 2 | 4
3 | 5 6

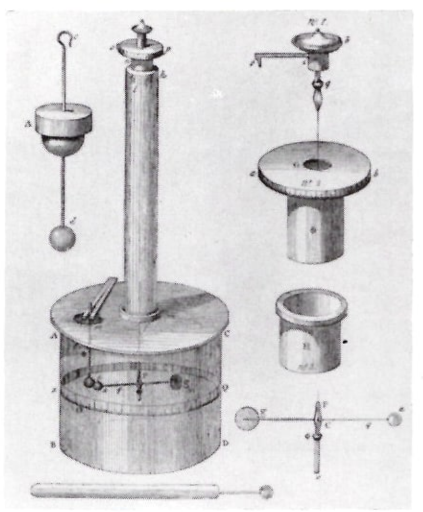
Tafeln 4 und 5

1 Universalmikroskop „George III." von ADAMS (um 1760) – **2** Luftteleskop aus dem 17. Jahrhundert – **3** Brennspiegel von

LAVOISIER zur Erzeugung hoher Temperaturen (um 1772). Der Linsendurchmesser betrug 1,5 m – **4** LAVOISIERS Präzisionswaage mit einer Wägegenauigkeit von 0,000 1 g – **5** Spiegelteleskop von HERSCHEL, mit dem er 1787 die Uranusmonde Oberon und Titanus entdeckte (Länge des Gerätes: 6 m) – **6** Die ersten Gewichte mit dezimaler Stückelung (um 1793).



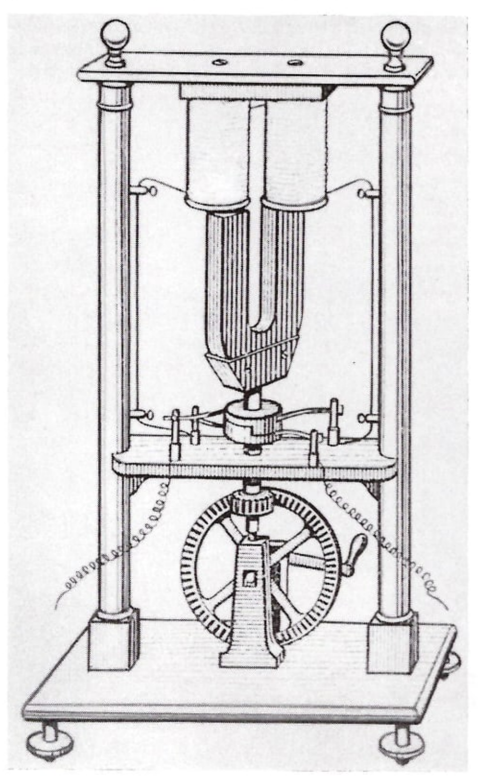


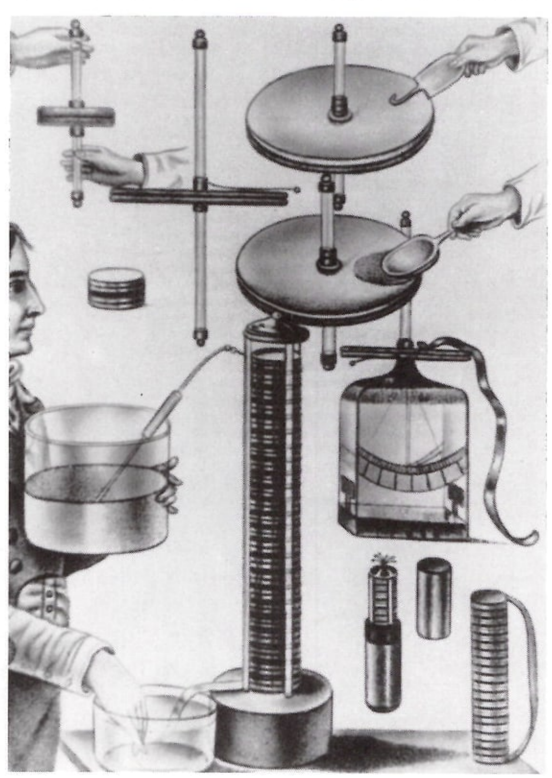


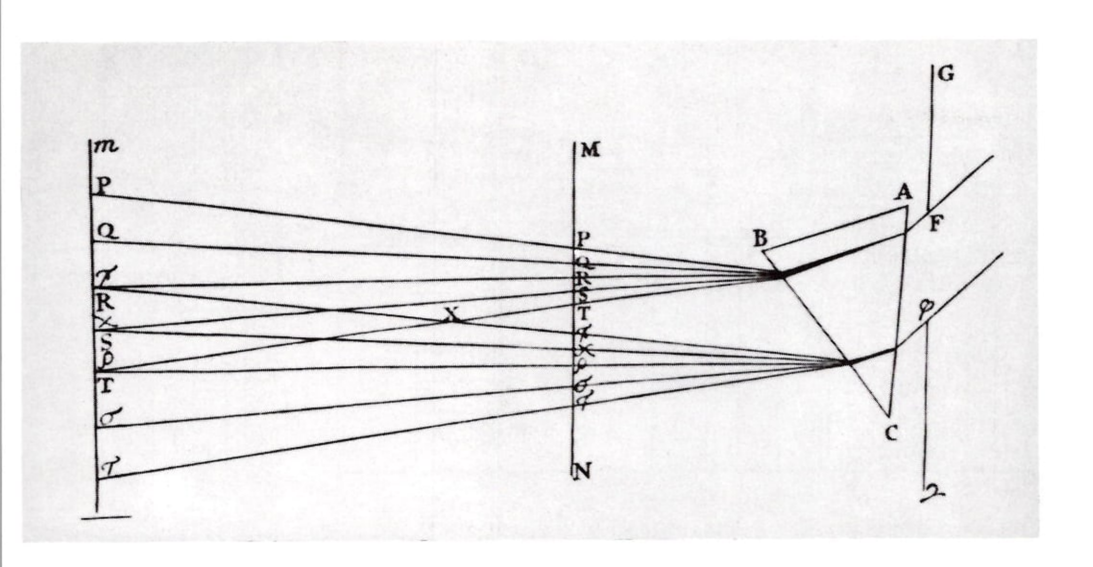


1 2 | 5
3 4 | 6 7

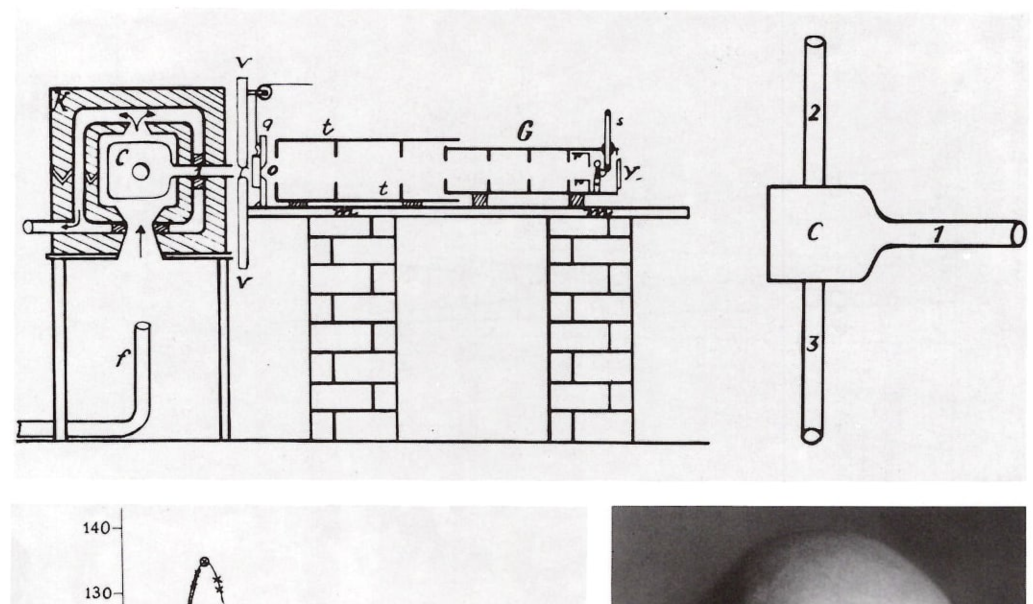
Tafeln 6 und 7
1 Drehwaage von Coulomb, mit der er das nach ihm benannte Gesetz über die Kraftwirkung zwischen elektrischen Ladungen fand – **2** Versuchsanordnung von Oersted für seine elektromagnetischen Experimente – **3** Magnetoelektrischer Gene-

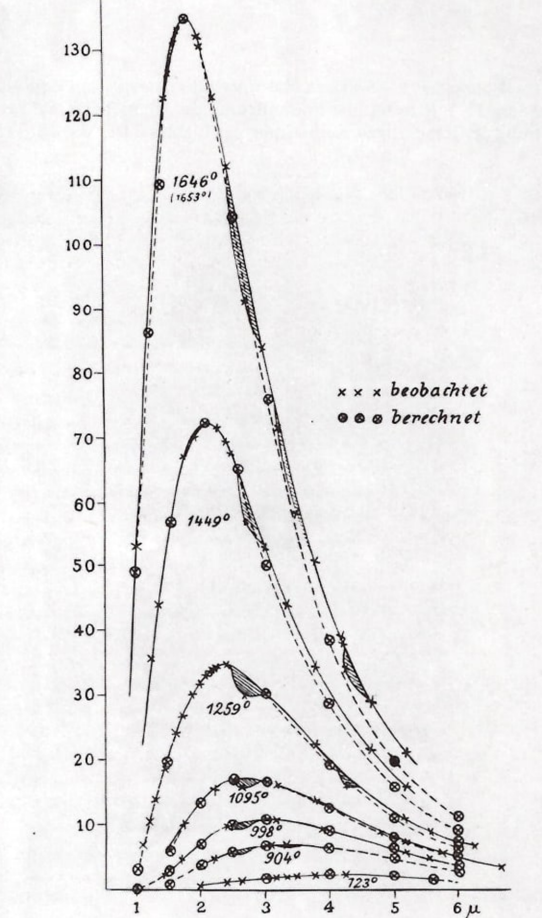


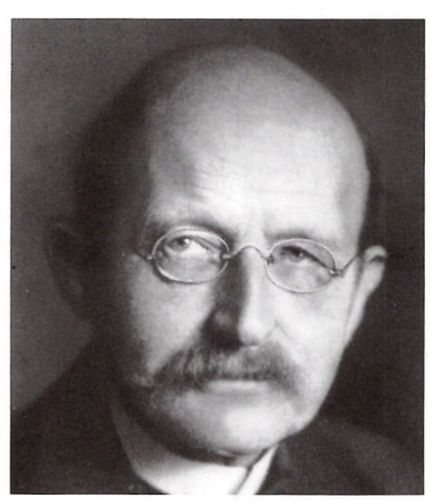


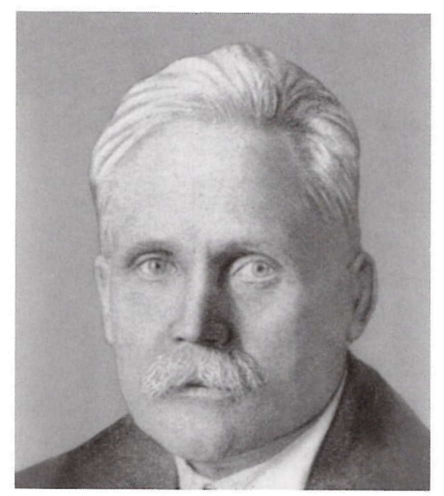


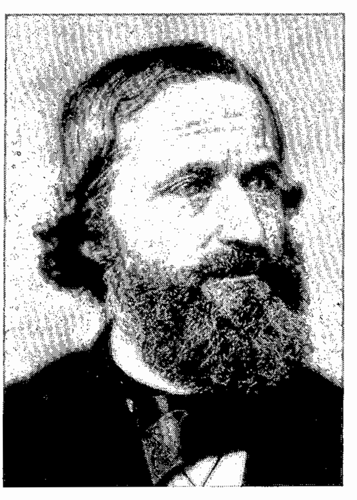
rator von AMPÈRE (um 1833) – **4** Darstellung der Versuche VOLTAS 1801 – **5** Skizze NEWTONS zur prismatischen Zerlegung des Lichtes – **6** Seite des Versuchsprotokolles von OHM vom 11. 1. 1826 mit der Formulierung des Ohmschen Gesetzes – **7** Seite aus dem Manuskript von HUYGENS mit der Erklärung der geradlinigen Ausbreitung des Lichtes durch das Huygenssche Prinzip.

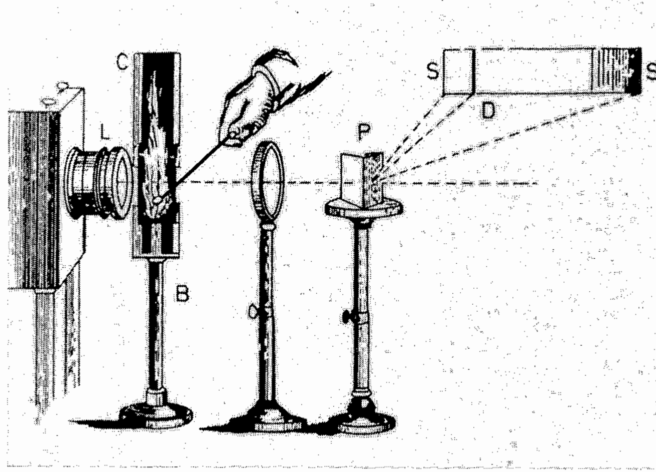

K
C
V
V
q
o
t
t
G
s
Y
f
2
C
1
3


140
130
120
110
100
90
80
70
60
50
40
30
20
10
1646°
(1653°)
1449°
1259°
1095°
998°
904°
723°
× × × beobachtet
⊗ ⊗ ⊗ berechnet
1
2
3
4
5
6 μ





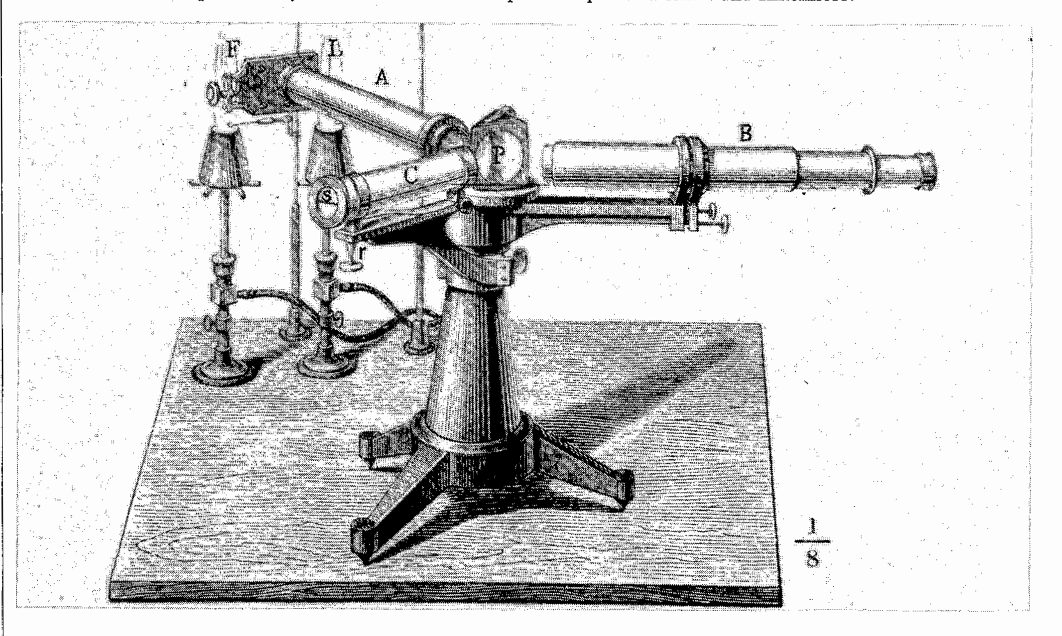




1 | 5 6
3 |
2 4 | 7

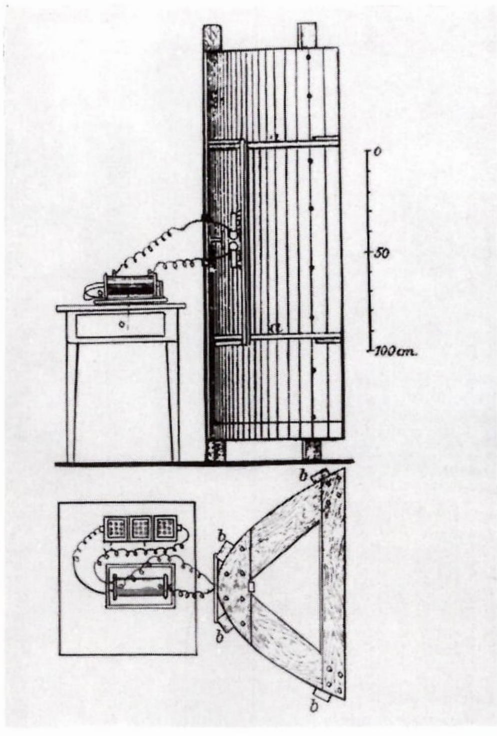
Tafeln 8 und 9

1 Schwarzer Strahler für Temperaturen über 600 °C nach LUMMER und PRINGSHEIM (1897). Der Hohlraum C (rechts, von oben gesehen) ist mit Schamotte ausgekleidet, die Heizung erfolgt von unten. Die Strahlung gelangt durch das Rohr 1 in die Tubusse t und G des Empfangssystems – 2 Meßergebnisse (ausgezogene Kurve) der spektralen Verteilung der Hohlraumstrahlung von LUMMER und PRINGSHEIM (1899). Die Unterschiede zu den aus der Wienschen Strahlungsgleichung erhaltenen gestrichelten Kurven veranlaßten PLANCK zur Aufstellung seiner Strahlungsgleichung – **3** MAX PLANCK (1858–1947) – **4** WILHELM WIEN (1864–1928) – **5** GUSTAV KIRCHHOFF (1824–1887) – **6** Versuchsanordnung von BUNSEN und KIRCHHOFF zur Spektralanalyse von Elementen. – **7** Spektroskop nach BUNSEN und KIRCHHOFF.



$\frac{1}{8}$



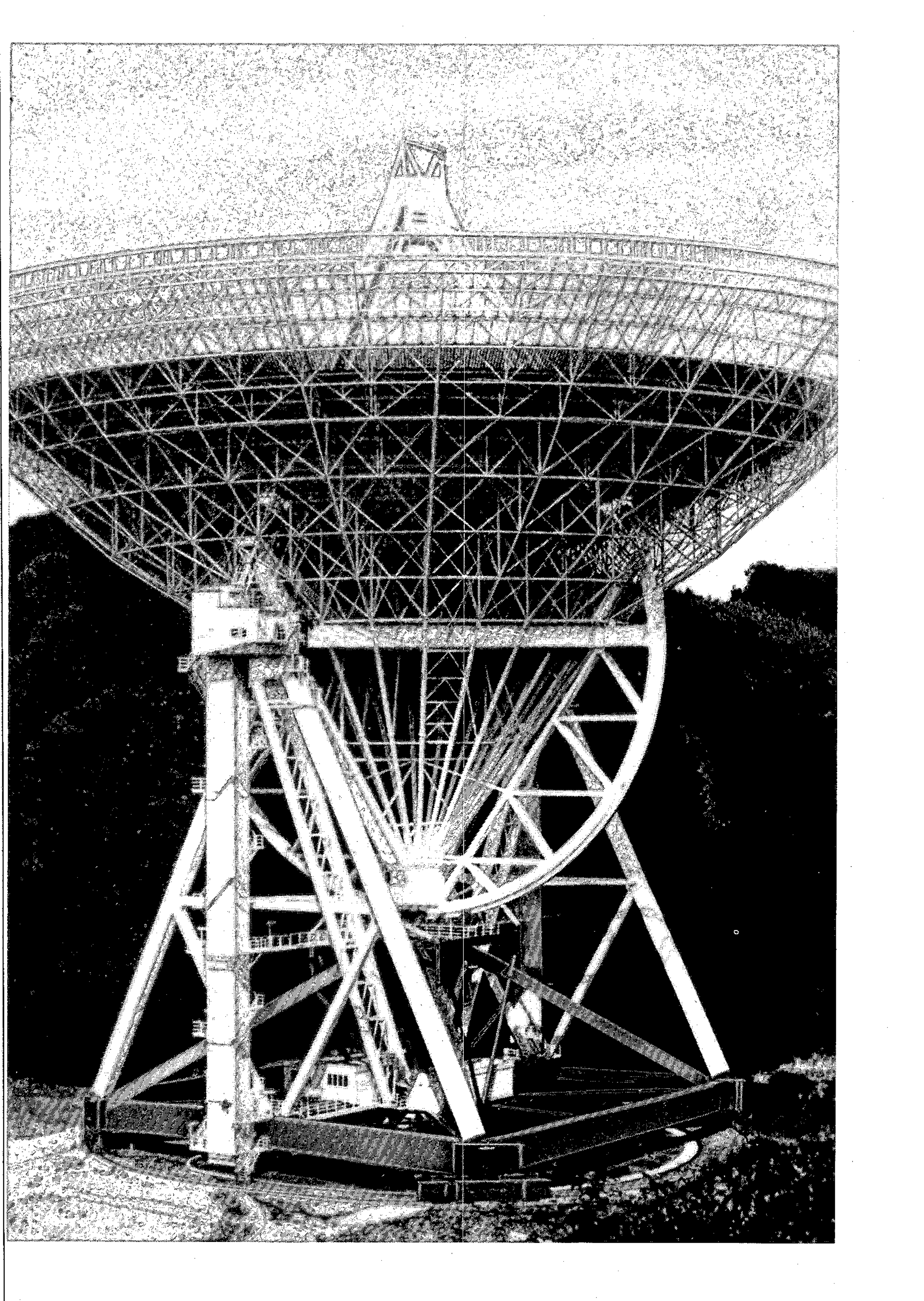


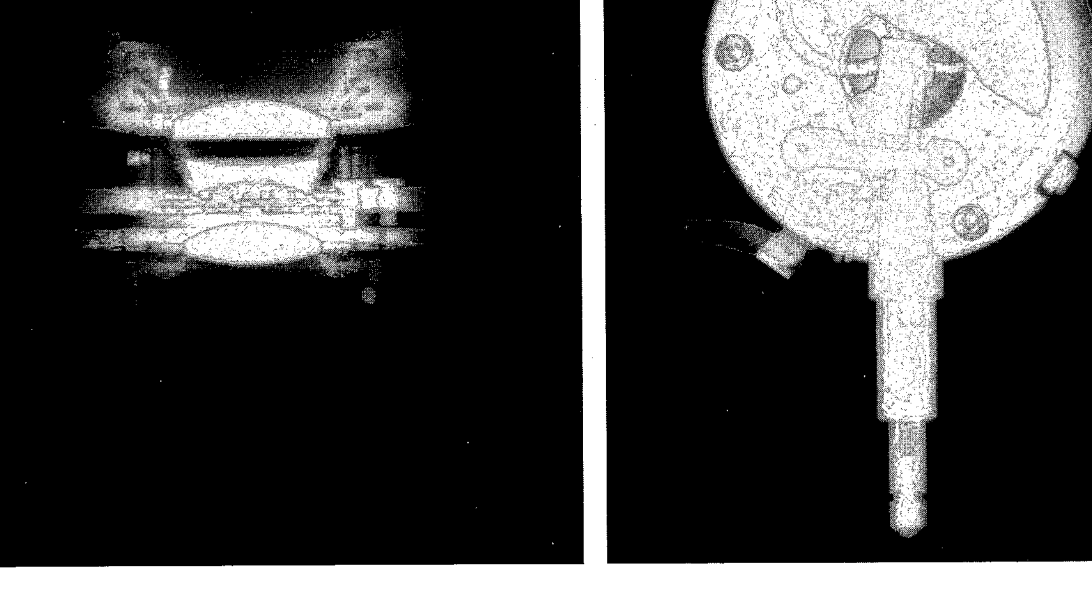
1 2 | 4
3

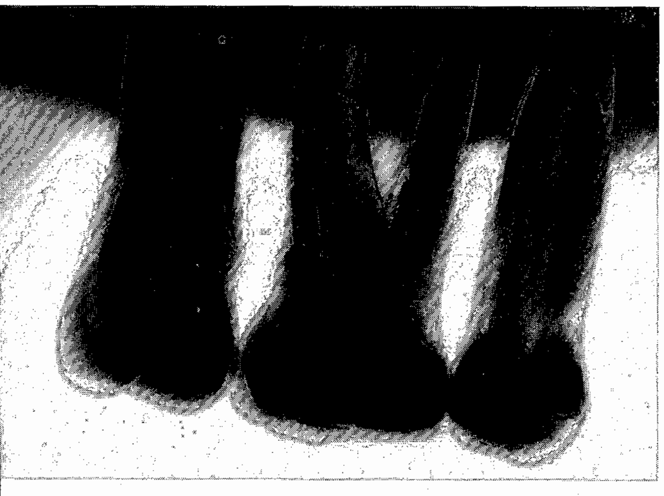


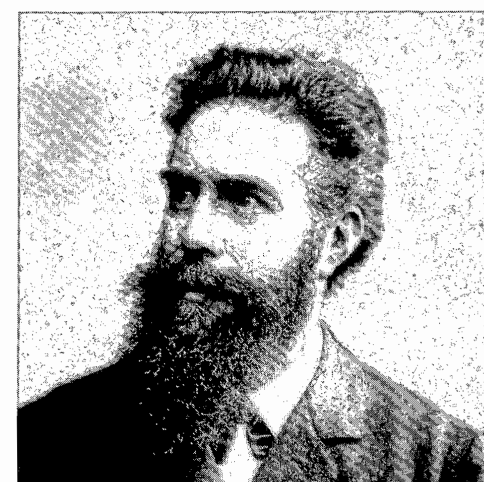
Tafeln 10 und 11

1 Von Hertz gebauter Resonator mit Funkenstrecke zum Nachweis elektromagnetischer Wellen im Dezimeterbereich – **2** Hertzscher Sender und Reflektor für Dezimalwellen (Zeichnung von Hertz) – **3** Heinrich Hertz (1857–1894) – **4** Das größte vollbewegliche Radioteleskop auf dem Effelsberg (Taunus), Spiegeldurchmesser 100 m.



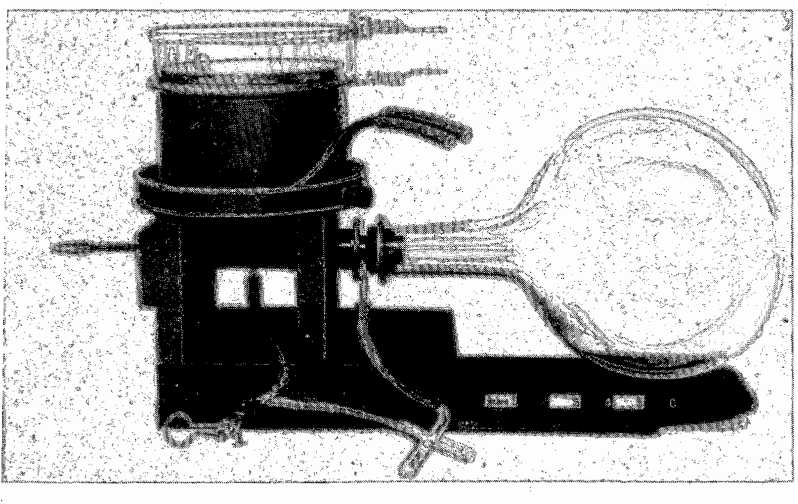


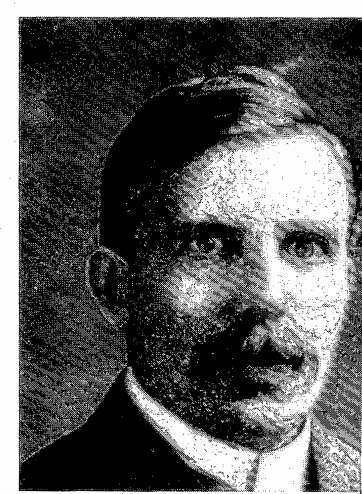




1 2 | 5 6
3 4 | 7

Tafeln 12 und 13
1/2 Röntgen-Aufnahme eines Kleinstmotors mit unterschiedlichen Kolbenstellungen – **3** Röntgen-Aufnahme vom Objektiv eines Fotoapparates – **4** Röntgen-Aufnahme einer Meßuhr – **5** Röntgen-Aufnahme von Zähnen mit erkennbaren Inlays – **6** WILHELM CONRAD RÖNTGEN (1845–1923) – **7** Röntgen-Aufnahme vom Becken eines gesunden zehnjährigen Kindes.

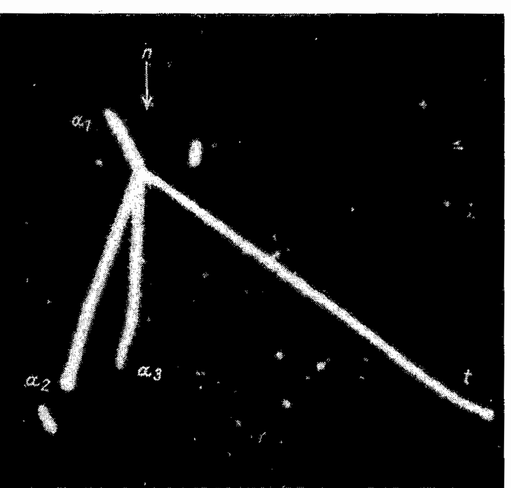


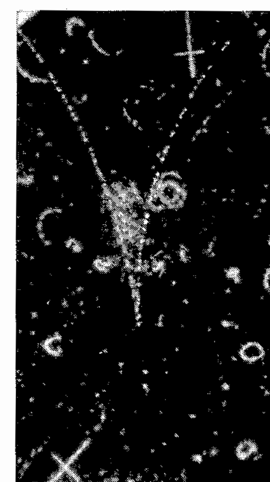


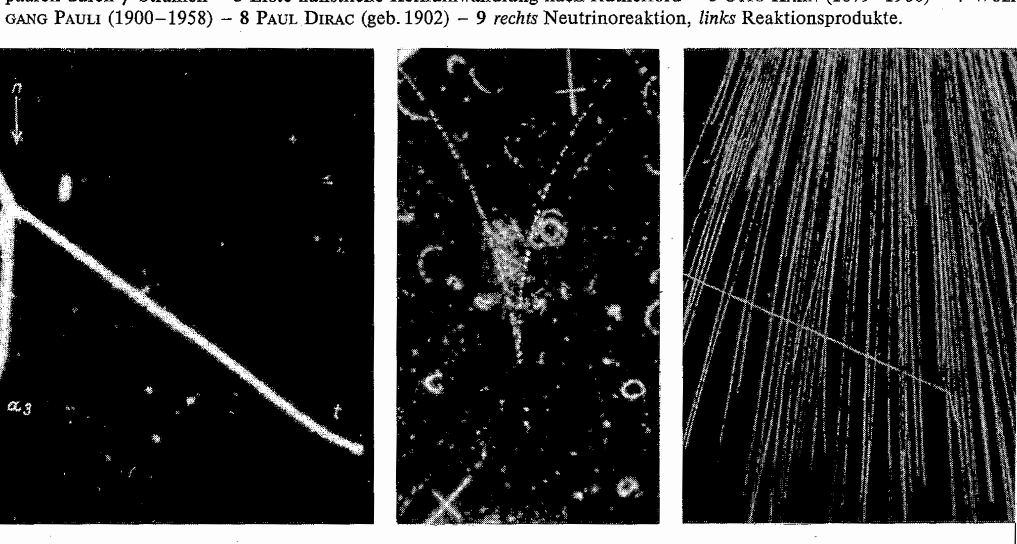
1 2 | 6 7 8
3 4 5 | 9

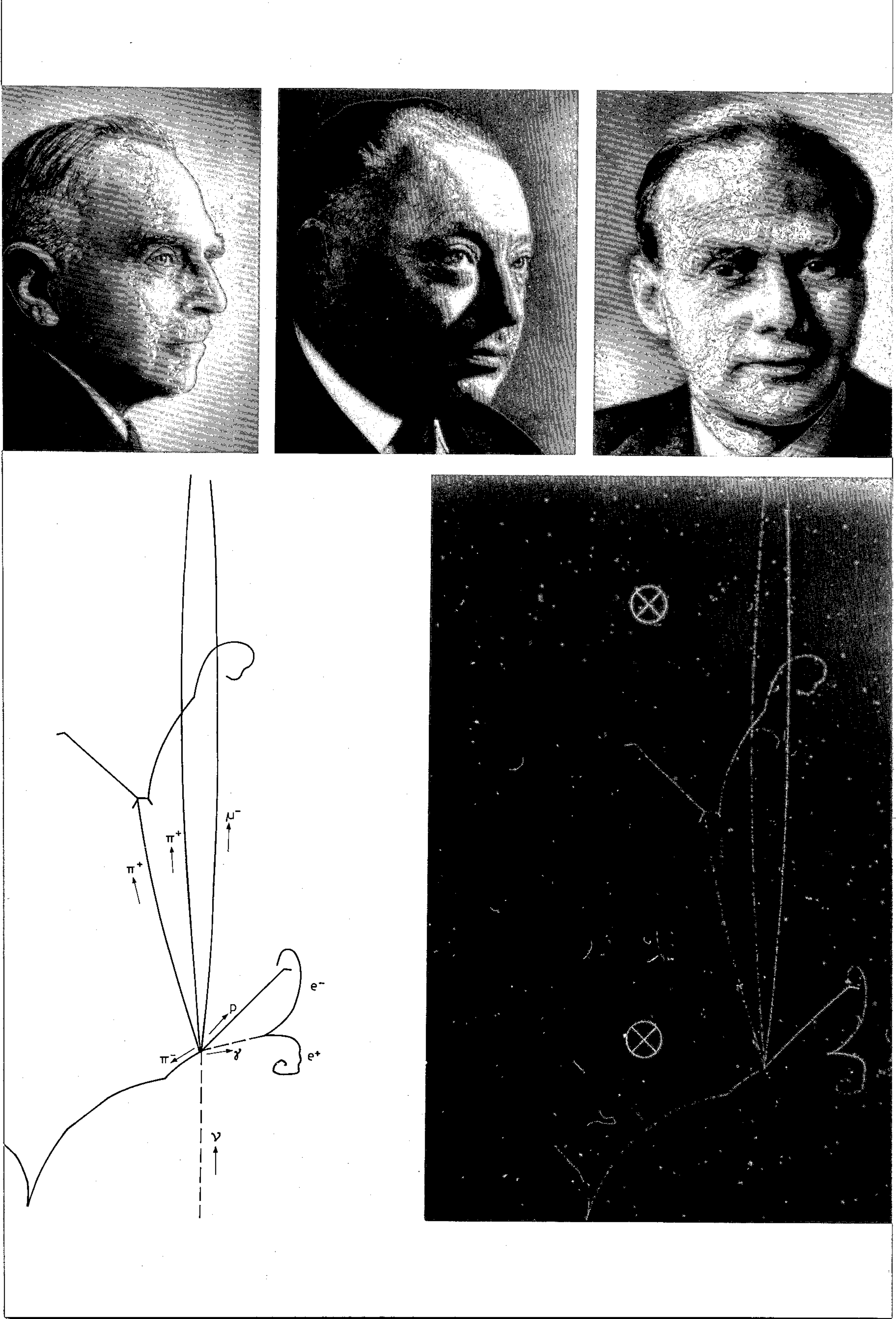
Tafeln 14 und 15

1 Erste Expansionsnebelkammer von WILSON (1912) – 2 ERNEST RUTHERFORD (1871–1937) – 3 Zertrümmerung eines Stickstoffkernes durch ein 15,7-MeV-Neutron, wobei drei α-Teilchen und ein Triton (t) entstehen – 4 Bildung von Elektronenpaaren durch γ-Strahlen – 5 Erste künstliche Kernumwandlung nach Rutherford – 6 OTTO HAHN (1879–1968) – 7 WOLFGANG PAULI (1900–1958) – 8 PAUL DIRAC (geb. 1902) – 9 *rechts* Neutrinoreaktion, *links* Reaktionsprodukte.







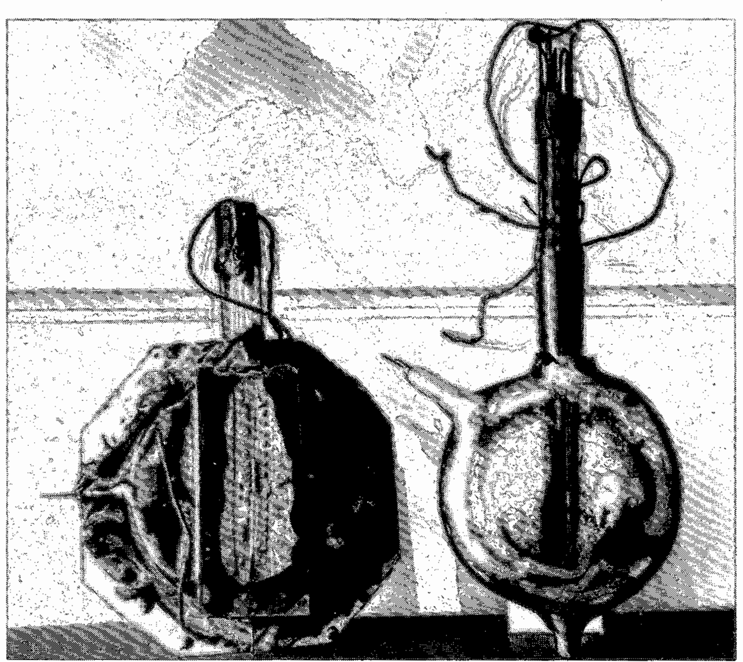
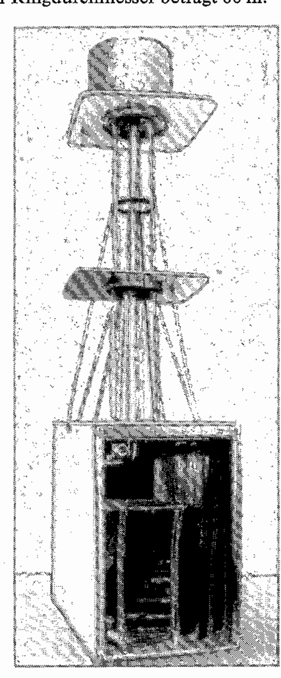


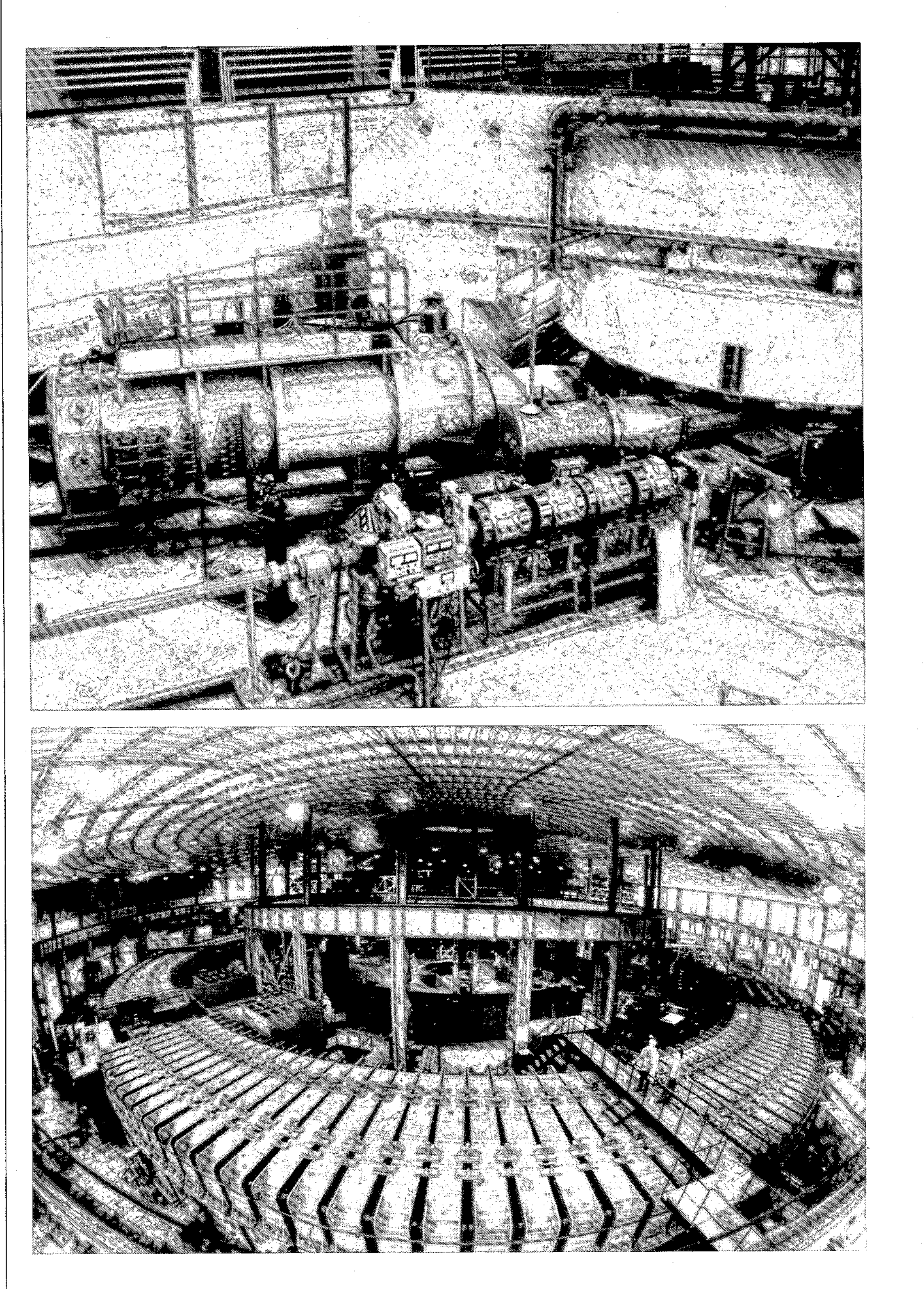
μ⁻
π⁺
π⁺
e⁻
p
π⁻
γ
e⁺
ν

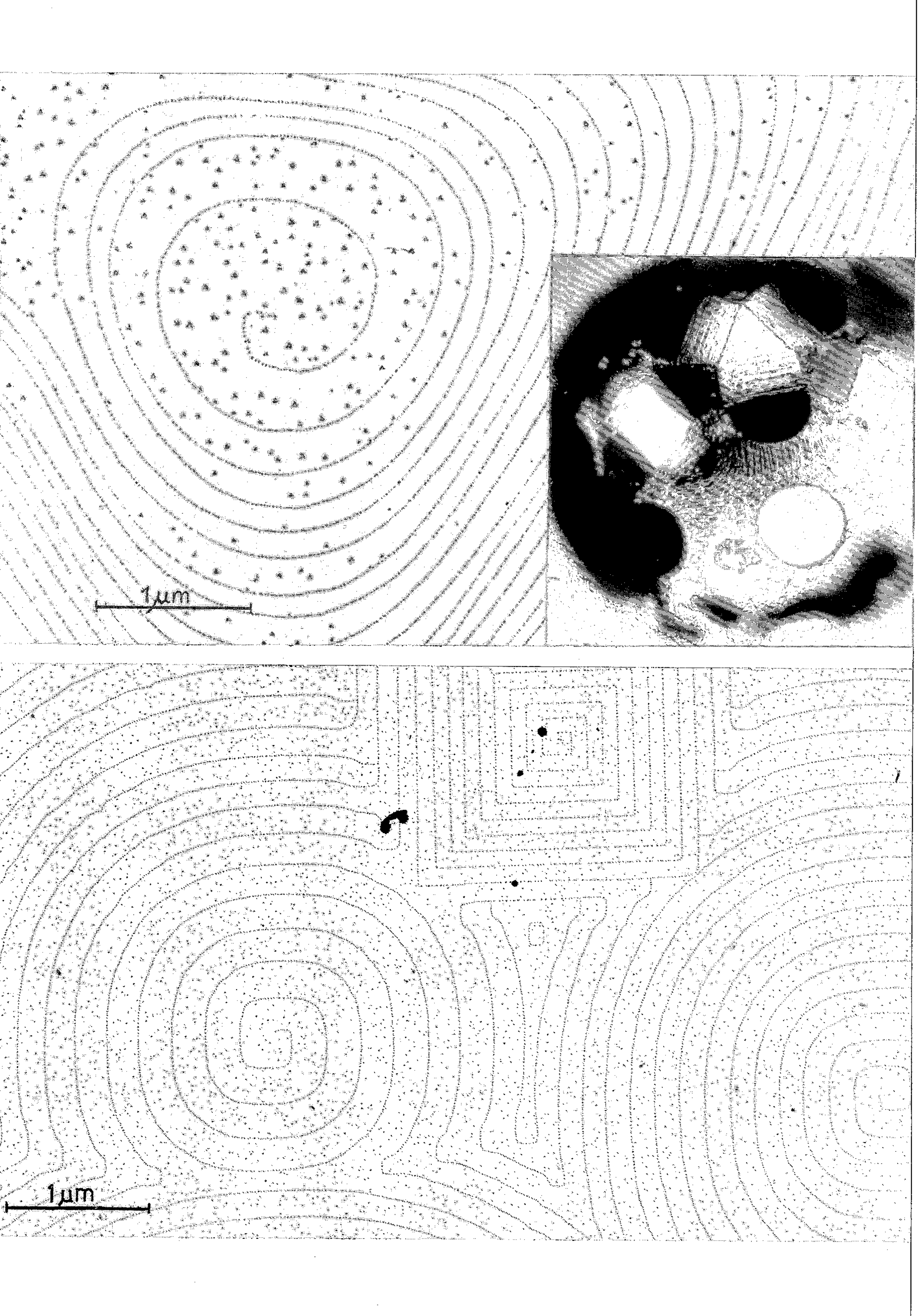



1 2 | 5
3 4 | 6

Tafeln 16 und 17
1 Ernest Orlando Lawrence (1901–1958) – 2 John Douglas Cockroft (1897–1967) – 3 Die ersten von Lawrence und Edlefsen konstruierten Modelle eines Zyklotrons – 4 Erste Beschleunigungsanlage von Cockroft und Walton zur künstlichen Kernzertrümmerung – 5 4-m-Zyklotron zur Beschleunigung schwerer Ionen – 6 Synchronphasotron in Dubna zur Beschleunigung von Protonen (10 GeV), Deuteronen (11 GeV) und α-Teilchen (22 GeV). Der Ringdurchmesser beträgt 60 m.

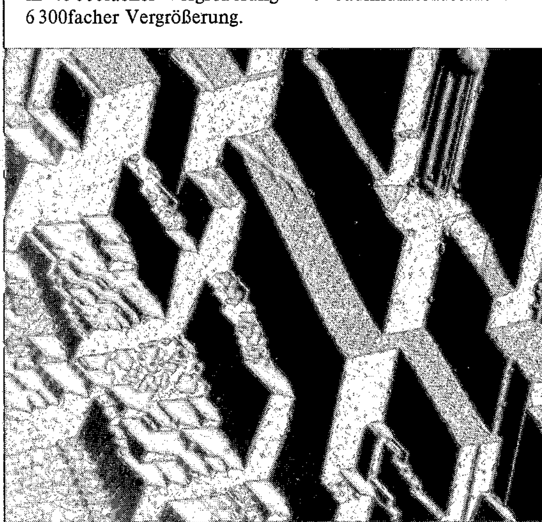



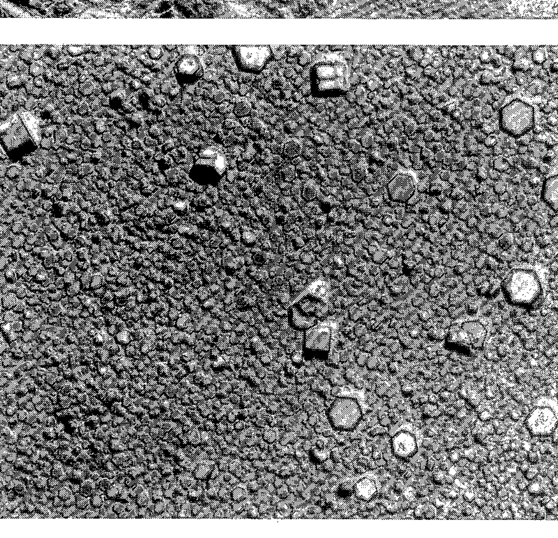


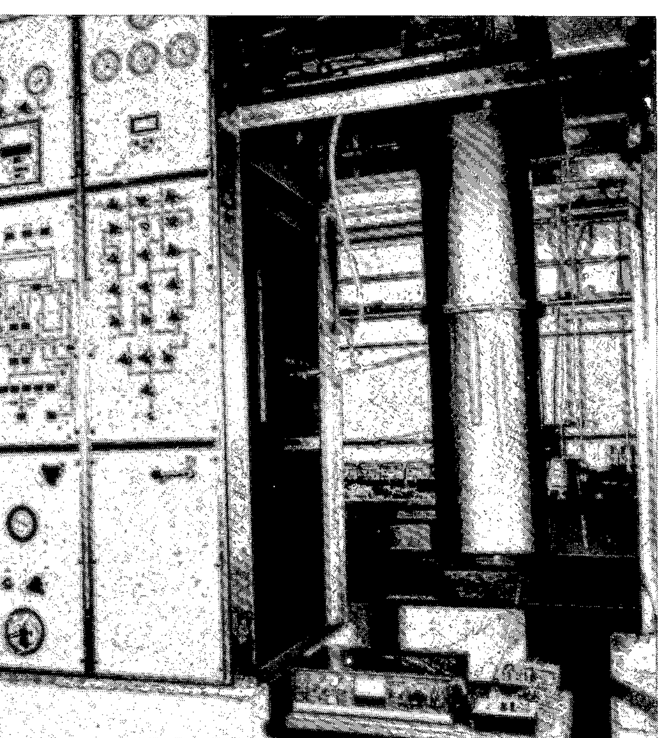

1 μm
1 μm

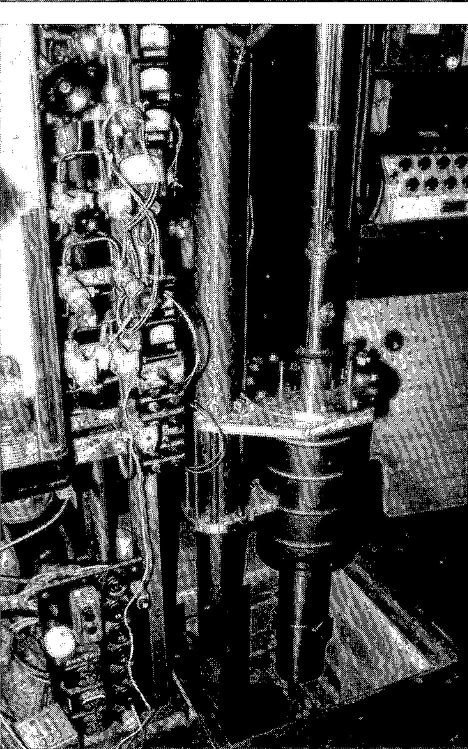
1 | 3
4
2 | 5 6

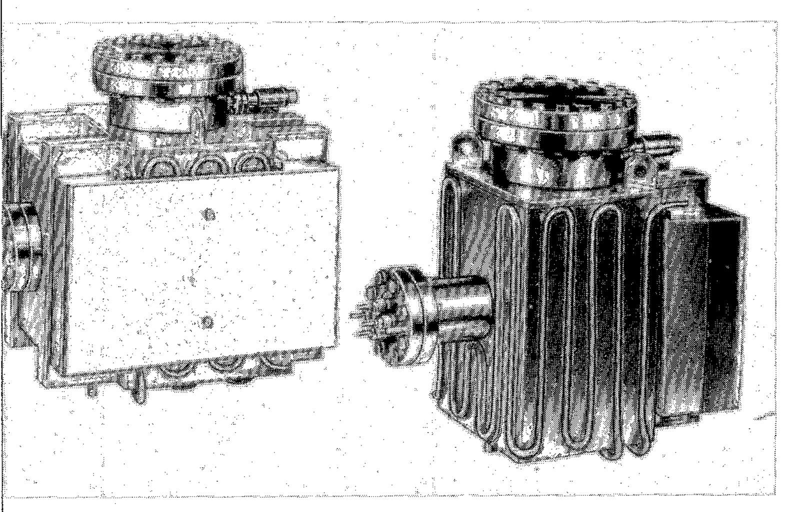
Tafeln 18 und 19
1 Durch Golddekoration abgebildete Wachstumsspirale auf der (111)-Fläche eines elektrolytisch gewachsenen kugelförmigen Silberkristalles – **2** Abdampfgruben an Versetzungsdurchstoßpunkten auf einer (100)-Fläche von Kochsalz. Die Stufenhöhe beträgt 0,28 bzw. 0,56 nm – **3** Silbersulfidwhiskerkristalle an Silberkontaktdrähten einer mikroelektronischen Schaltung – **4** Elektronenmikroskopische Aufnahme einer Glasoberfläche – **5** Oberfläche eines Platineinkristalles in 40 000facher Vergrößerung – **6** Cadmiumoxidschicht in 6 300facher Vergrößerung.

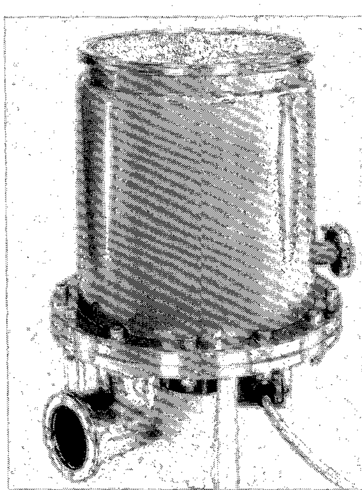








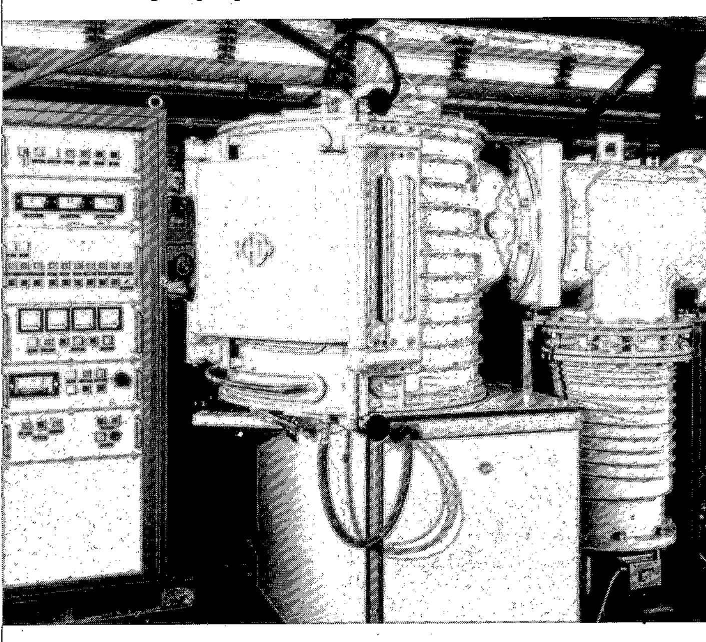


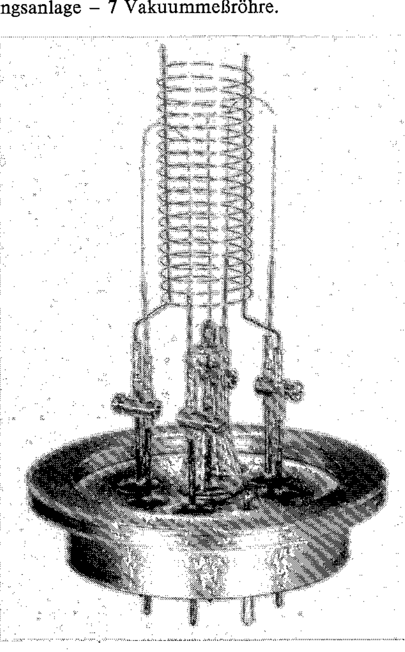


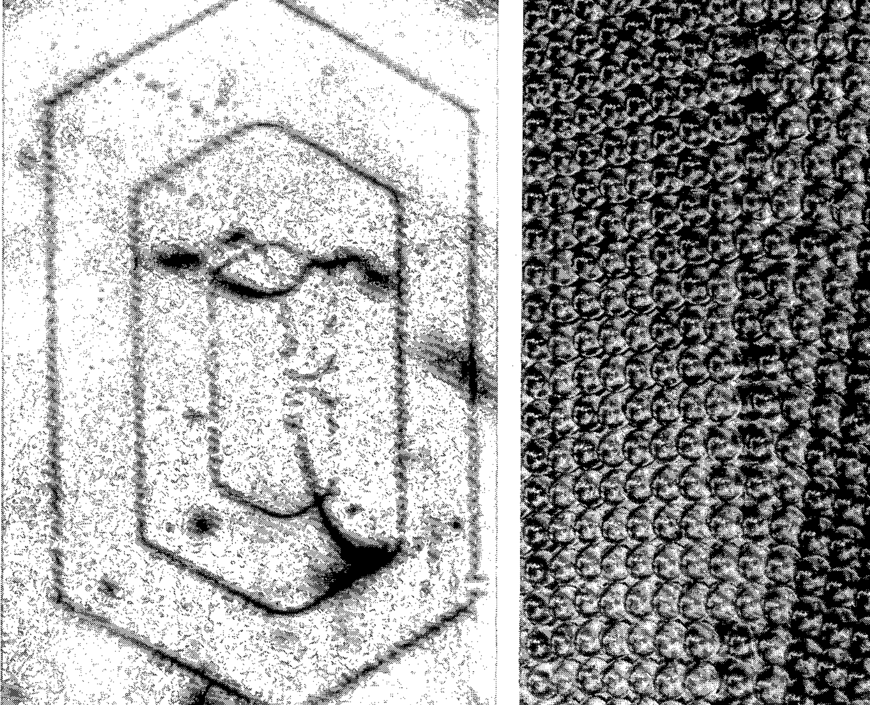
1 | 4 5
2 3 | 6 7

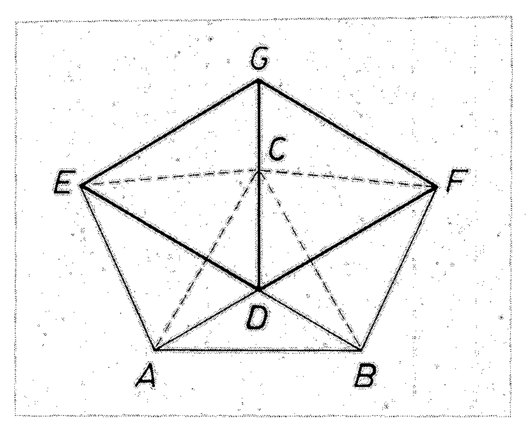
Tafeln 20 und 21

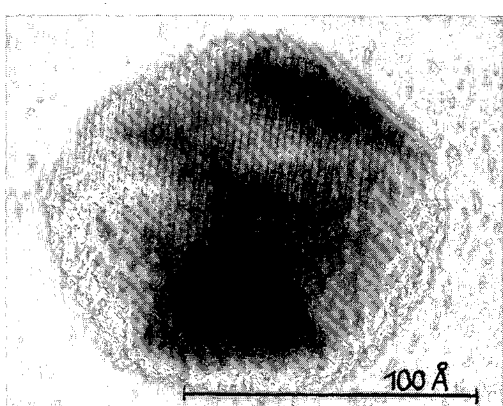
1 Anlage zur Heliumverflüssigung mit einer Leistung von 10 l/h – 2 $^{3}He/^{4}He$-Entmischungskältemaschine, mit der im Dauerbetrieb eine Temperatur von 0,02 K gehalten werden kann – 3 Kryostat, mit dem Temperaturen bis zu 0,4 K erreicht werden – 4 Ionengetterpumpen – 5 Turbomolekularpumpe – 6 Hochvakuumbedampfungsanlage – 7 Vakuummeßröhre.





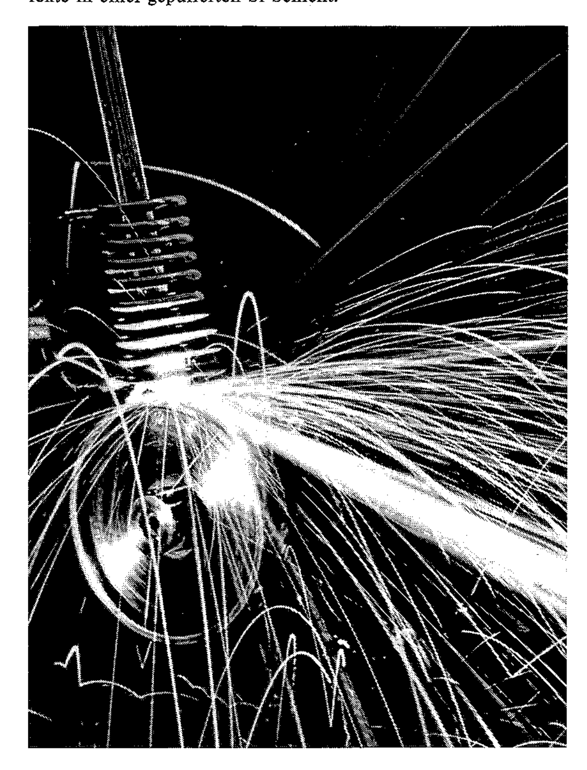

101
001
B
C
C
113
D
A
011
B
C

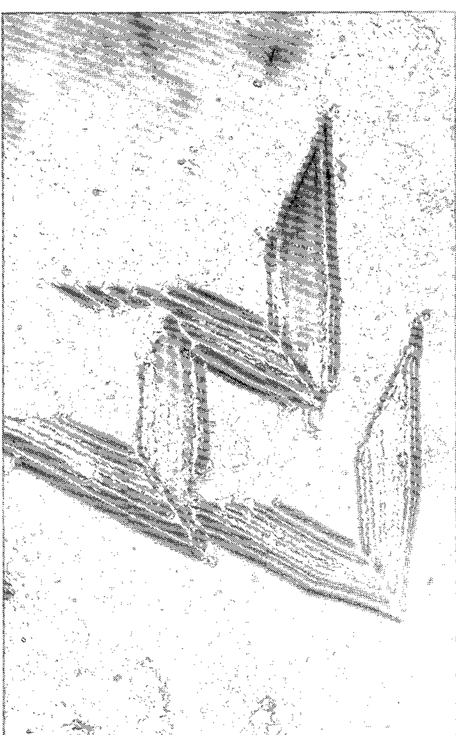


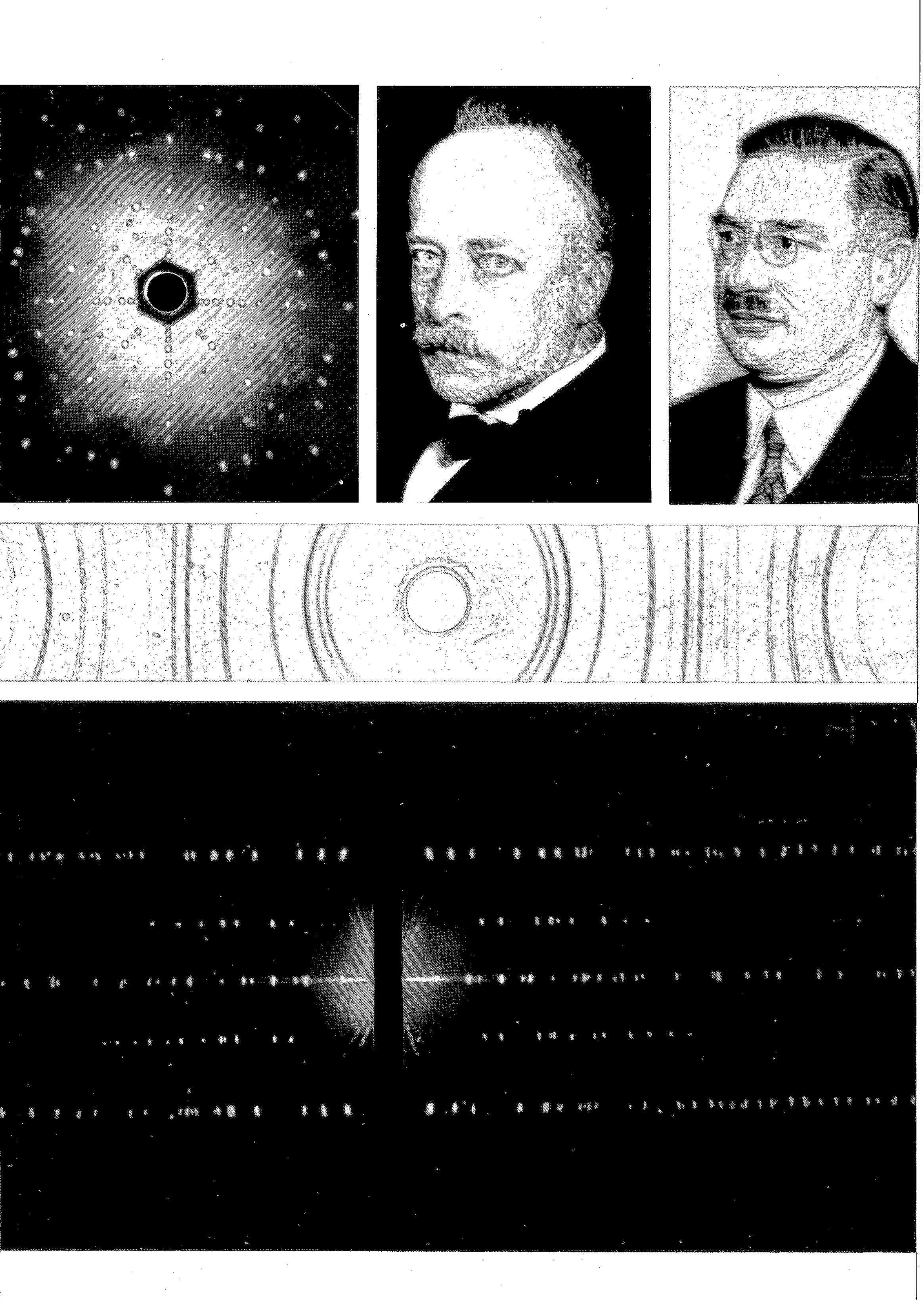


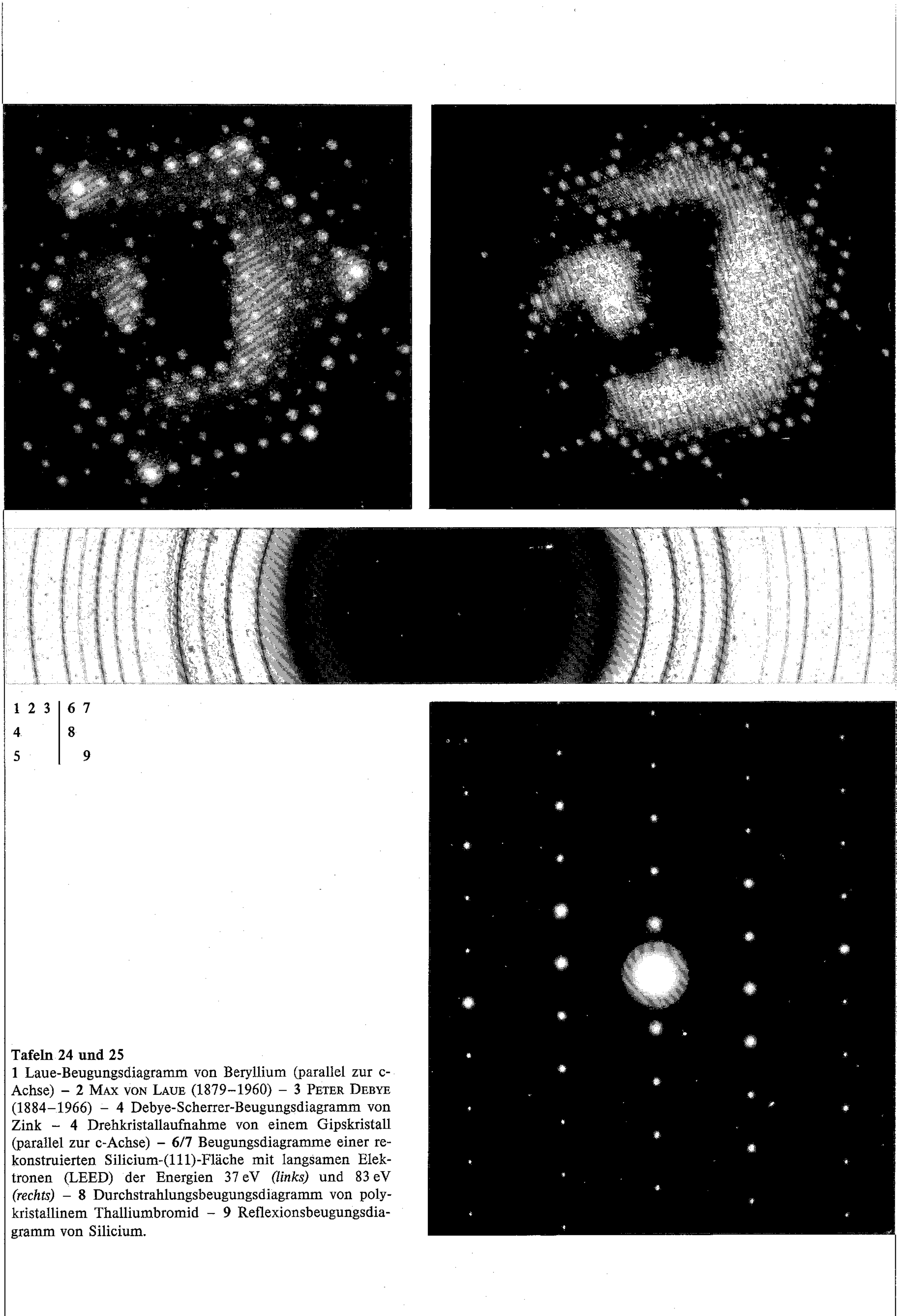
1 | 4
2 3 | 5 6

Tafeln 22 und 23
1 Feldionenmikroskopische Abbildung einer Platinspitze mit Kristallbaufehlern – **2** Infrarotfoto einer Frank-Read-Versetzungsquelle (nach DASH) – **3** Kugelmodell einer Kleinwinkelkorngrenze – **4** Elektronenmikroskopische Durchstrahlungsaufnahme eines Goldteilchens mit pentagonaler Form (Skizze links). Die parallel zum Elektronenstrahl liegenden Atomebenen sind als parallele Streifen sichtbar – **5** Herstellung amorpher Metallfolien. Das in einem Quarzrohr befindliche Metall wird induktiv geschmolzen und durch eine Düse gegen die Oberfläche eines mit 3 000 min^{-1} rotierenden Kupferzylinders gepreßt , wo es in weniger als 10^{-3} s zu einer bis μm dicken Folie erstarrt (helles Band) – **6** Pyramidenförmige Defekte in einer gepufferten Si-Schicht.



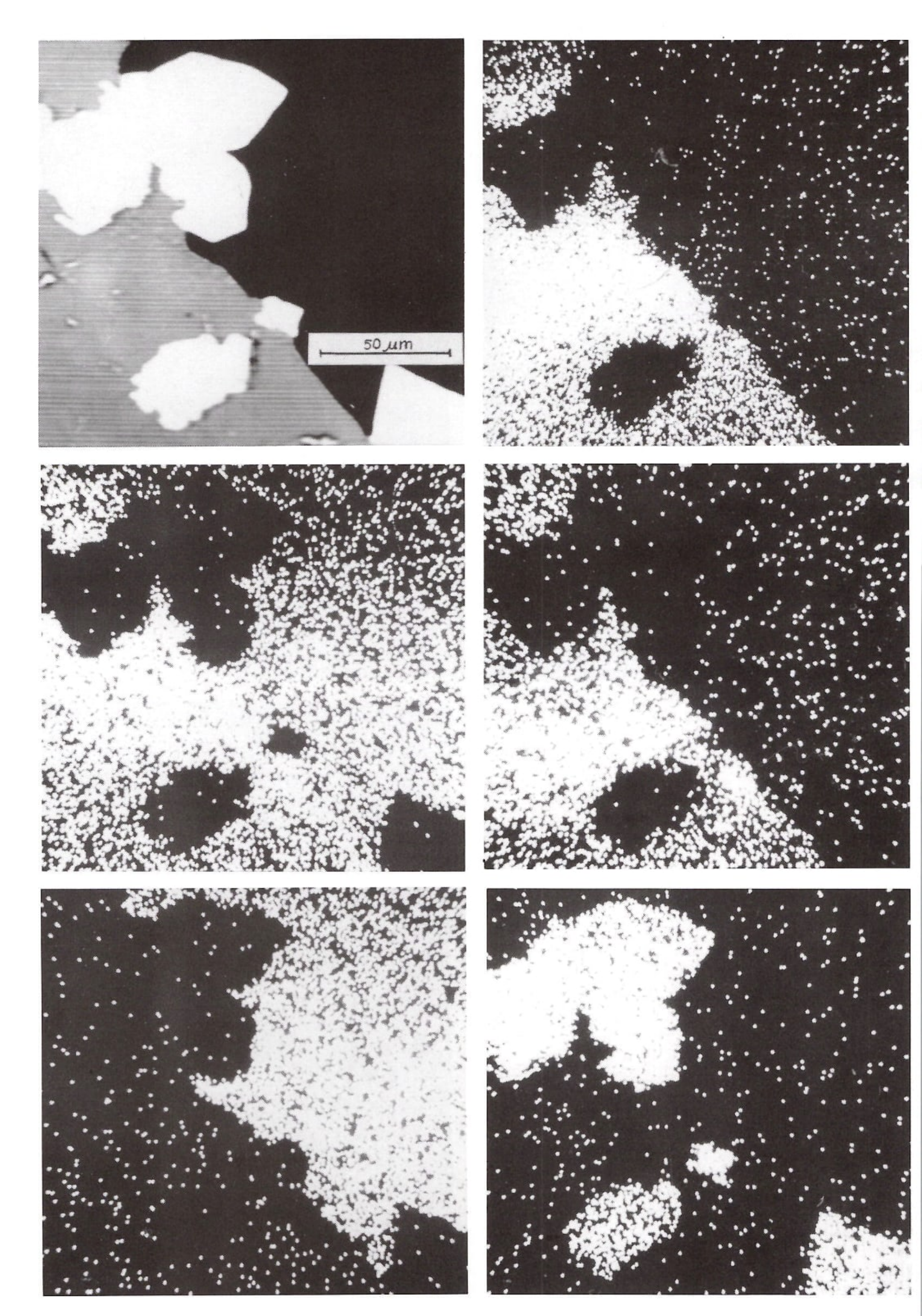


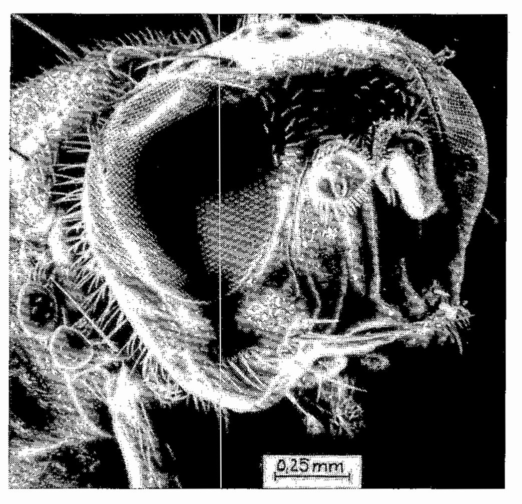


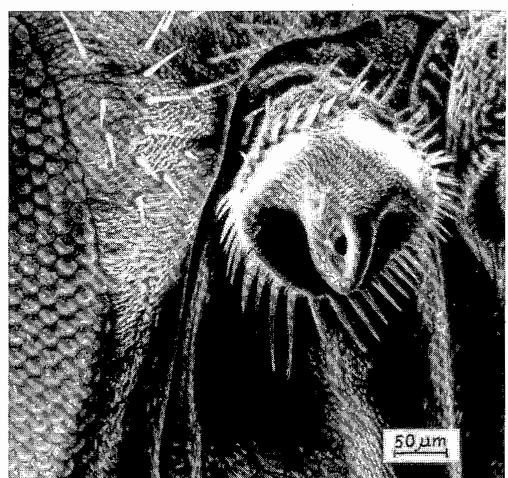


1 2 3	6 7
4	8
5	9

Tafeln 24 und 25
1 Laue-Beugungsdiagramm von Beryllium (parallel zur c-Achse) – **2** MAX VON LAUE (1879–1960) – **3** PETER DEBYE (1884–1966) – **4** Debye-Scherrer-Beugungsdiagramm von Zink – **4** Drehkristallaufnahme von einem Gipskristall (parallel zur c-Achse) – **6/7** Beugungsdiagramme einer rekonstruierten Silicium-(111)-Fläche mit langsamen Elektronen (LEED) der Energien 37 eV *(links)* und 83 eV *(rechts)* – **8** Durchstrahlungsbeugungsdiagramm von polykristallinem Thalliumbromid – **9** Reflexionsbeugungsdiagramm von Silicium.


50 µm

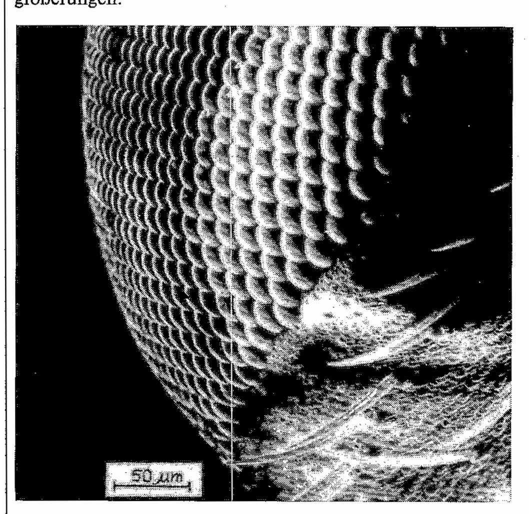




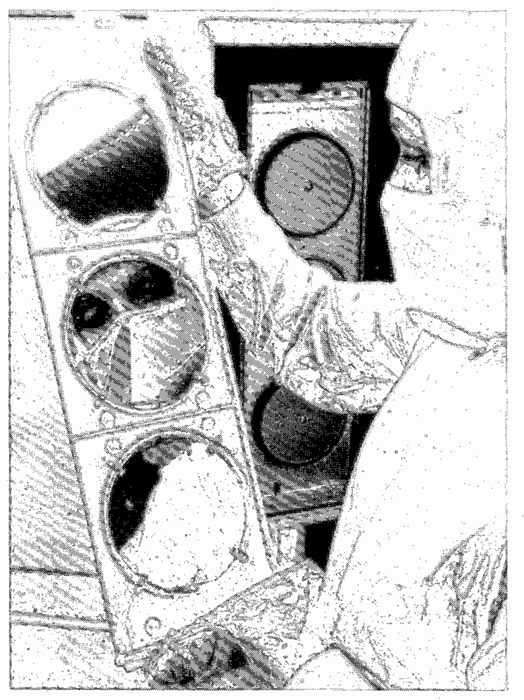
1 2	7 8
3 4	
5 6	9 10

Tafeln 26 und 27

1–6 Mit Mikrosondeanalysator erhaltene Rasterbilder einer Erzprobe mit den Bestandteilen Quarz, Bleiglanz, und eisenhaltiges Zinksulfid – **1** Probenstrombild – **2** Eisenstrahlung – **3** Schwefelstrahlung – **4** Zinkstrahlung – **5** Bleistrahlung – **6** Siliciumstrahlung – **7–10** Rasterelektronenmikroskopische Aufnahmen vom Kopf einer Stubenfliege mit Ausschnittvergrößerungen.



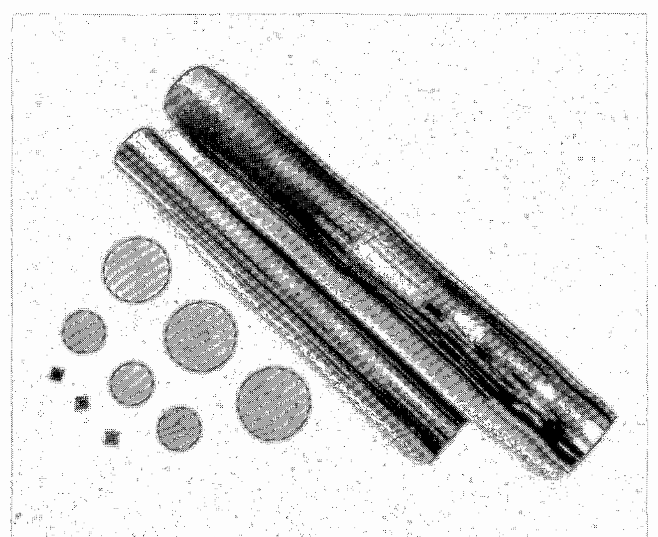


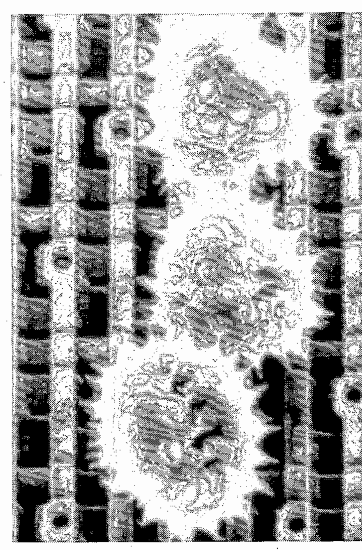


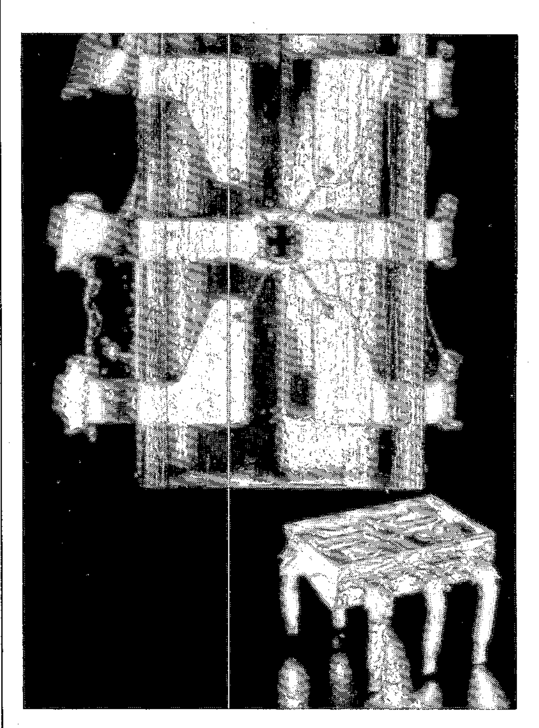


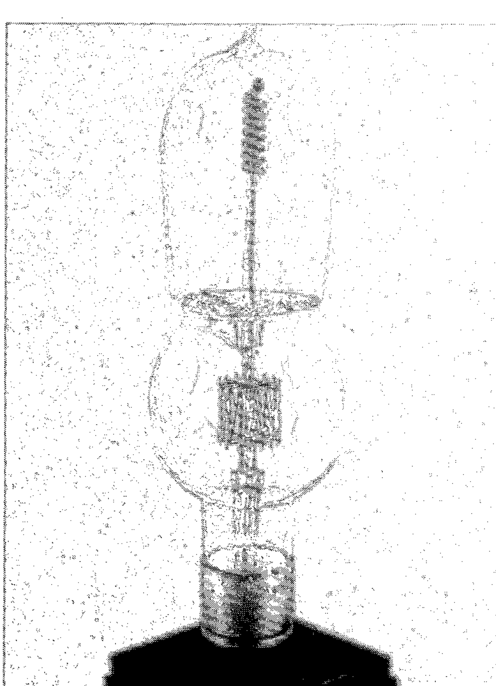
1 2 | 5 6
3 4 | 7 8

Tafeln 28 und 29
1 Siliciumeinkristallscheiben (Wafer) für Speicherbausteine (Durchmesser 5 Zoll) – **2** Elektronische Prüfung der Chips einer Siliciumscheibe mit einem Vielfachsondentester – **3** Siliciumeinkristalle in verschiedenen Bearbeitungsstadien –

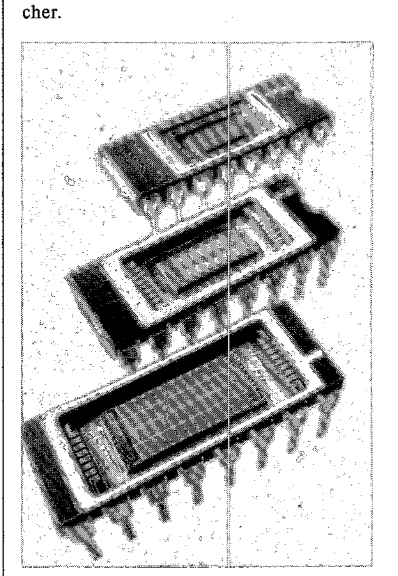


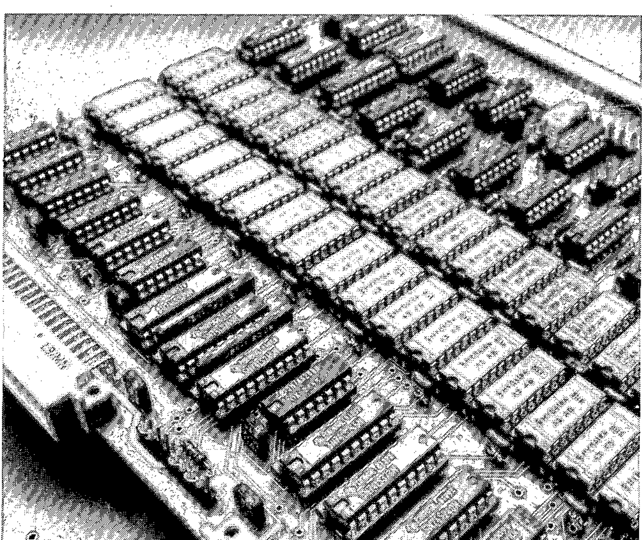


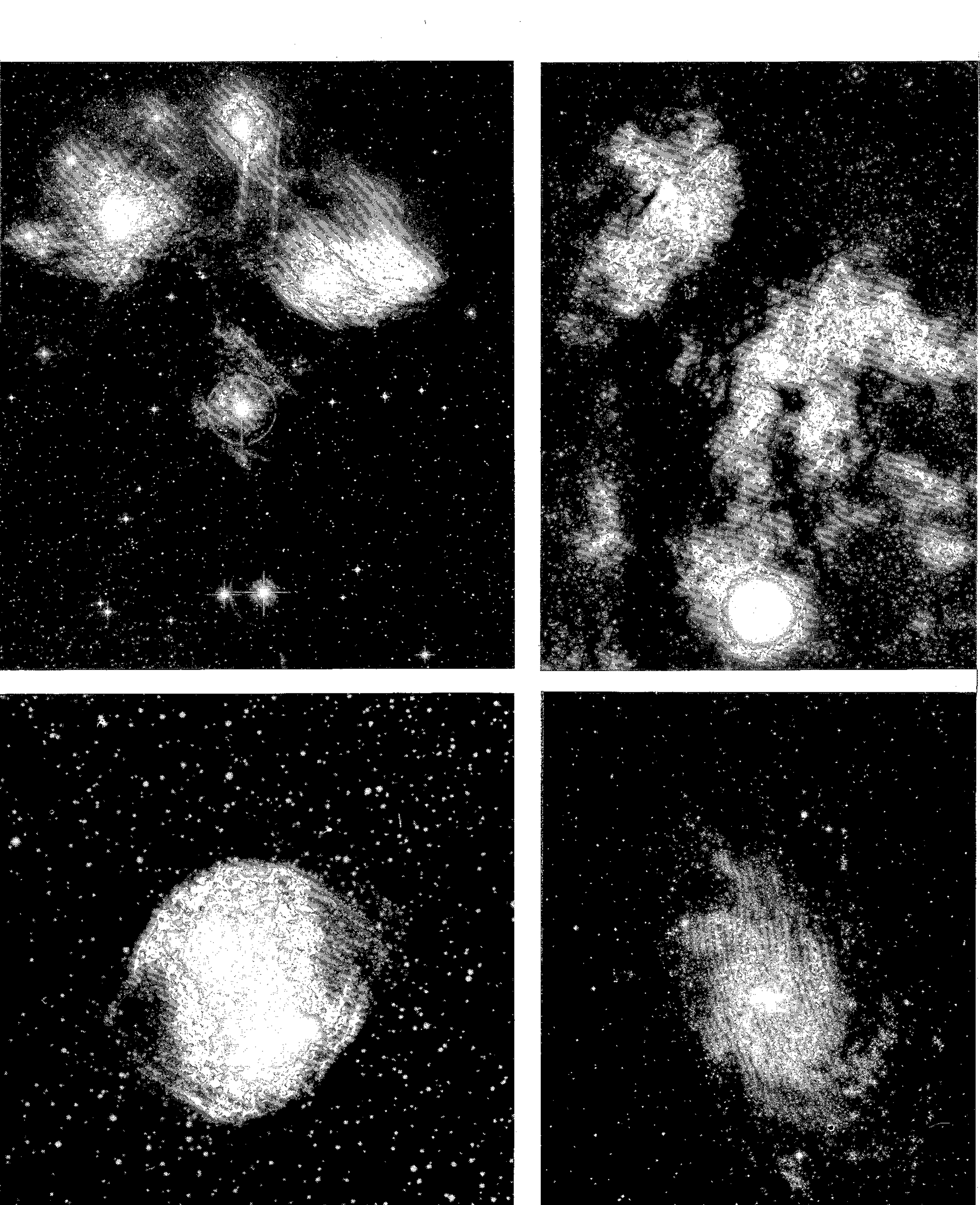




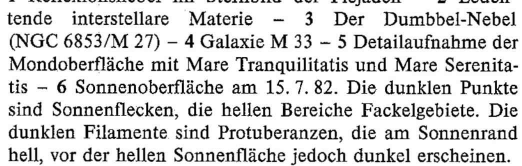
4 Leiterbahnen eines 256-k-Chips mit 20 μm großen Blütenstaubpollen zur Verdeutlichung, wie empfindlich diese Bausteine gegen Verunreinigungen sind – **5** Fotodiode, deren vier aktive Teilflächen im Diagonalschnittpunkt einen Abstand von nur 12 μm haben. Dadurch können kleinste Abweichungen des auftretenden Lichtes registriert werden – **6** Älteste Elektronenröhre, die 1906 von BARKHAUSEN (1881–1956) zum Patent angemeldet wurde – **7** von oben nach unten: 16-k-Speicher. Letzterer hat eine Integrationsdichte von 10 000 Transistoren/mm² – **8** Mikrocomputerbaugruppe mit 256-kByte-Speicher.





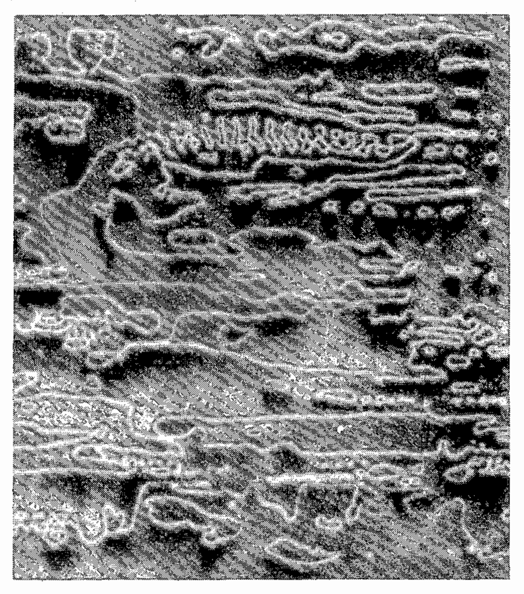


1 2 | 5
3 4 | 6



Tafeln 30 und 31

1 Reflexionsnebel im Sternbild der Plejaden – 2 Leuchtende interstellare Materie – 3 Der Dumbbel-Nebel (NGC 6853/M 27) – 4 Galaxie M 33 – 5 Detailaufnahme der Mondoberfläche mit Mare Tranquilitatis und Mare Serenitatis – 6 Sonnenoberfläche am 15. 7. 82. Die dunklen Punkte sind Sonnenflecken, die hellen Bereiche Fackelgebiete. Die dunklen Filamente sind Protuberanzen, die am Sonnenrand hell, vor der hellen Sonnenfläche jedoch dunkel erscheinen.



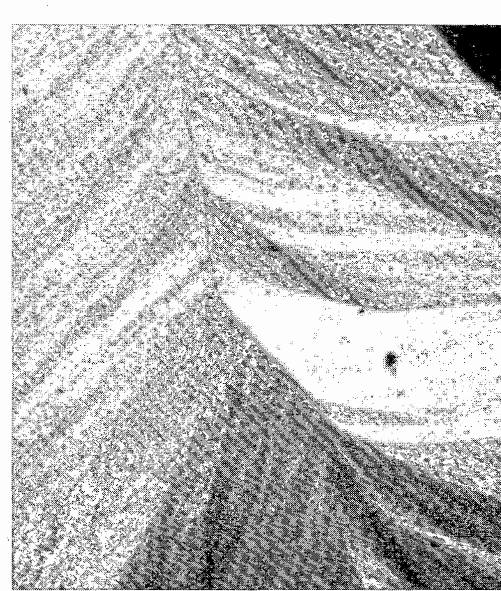


1 2
3 4

Tafel 32

1 Rasterelektronische Aufnahme der ferroelektrischen Domänen eines GASH-Einkristalles (Guandinaluminiumhexahydrat) – 2 Elektronenmikroskopische Aufnahme der Chromiumoxidschicht eines Videomagnetbandes (10 000fache Vergrößerung) – 3 Weißsche Bezirke in einem Einkristall eines siliciumhaltigen Eisens (die Pfeile geben die Magnetisierungsrichtung an.) – 4 Elektronenmikroskopische Aufnahme der Magnetstruktur in einer 60 nm dicken Nickeleisenschicht (5 000fache Vergrößerung).





Tafel 33
Periodensystem der Elemente

1a	2a	3b	4b	5b	6b	7b	8			1b	2b	3a	4a	5a	6a	7a	0	Gruppe / Periode
1 H 1,0079																	2 He 4,0026	1
3 Li 6,941	4 Be 9,0122											5 B 10,81	6 C 12,011	7 N 14,0067	8 O 15,9994	9 F 18,9984	10 Ne 20,179	2
11 Na 22,9898	12 Mg 24,305											13 Al 26,9815	14 Si 28,0855	15 P 30,9738	16 S 32,06	17 Cl 35,453	18 Ar 39,948	3
19 K 39,0983	20 Ca 40,08	21 Sc 44,9559	22 Ti 47,90	23 V 50,9414	24 Cr 51,996	25 Mn 54,9380	26 Fe 55,847	27 Co 58,9332	28 Ni 58,70	29 Cu 63,546	30 Zn 65,38	31 Ga 69,72	32 Ge 72,59	33 As 74,9216	34 Se 78,96	35 Br 79,904	36 Kr 83,80	4
37 Rb 85,4678	38 Sr 87,62	39 Y 88,9059	40 Zr 91,22	41 Nb 92,9064	42 Mo 95,94	43 Tc (97)	44 Ru 101,07	45 Rh 102,9055	46 Pd 106,4	47 Ag 107,868	48 Cd 112,41	49 In 114,82	50 Sn 118,69	51 Sb 121,75	52 Te 127,60	53 I 126,9045	54 Xe 131,29	5
55 Cs 132,9054	56 Ba 137,33	57* La 138,9055	72 Hf 178,49	73 Ta 180,9479	74 W 183,85	75 Re 186,207	76 Os 190,2	77 Ir 192,22	78 Pt 195,09	79 Au 196,9665	80 Hg 200,59	81 Tl 204,37	82 Pb 207,2	83 Bi 208,9804	84 Po (209)	85 At (210)	86 Rn (222)	6
87 Fr (223)	88 Ra 226,0254	89** Ac 227,0278	104 Ku (261)	105 Ns (262)	106 Unh (263)	107 Uns												7

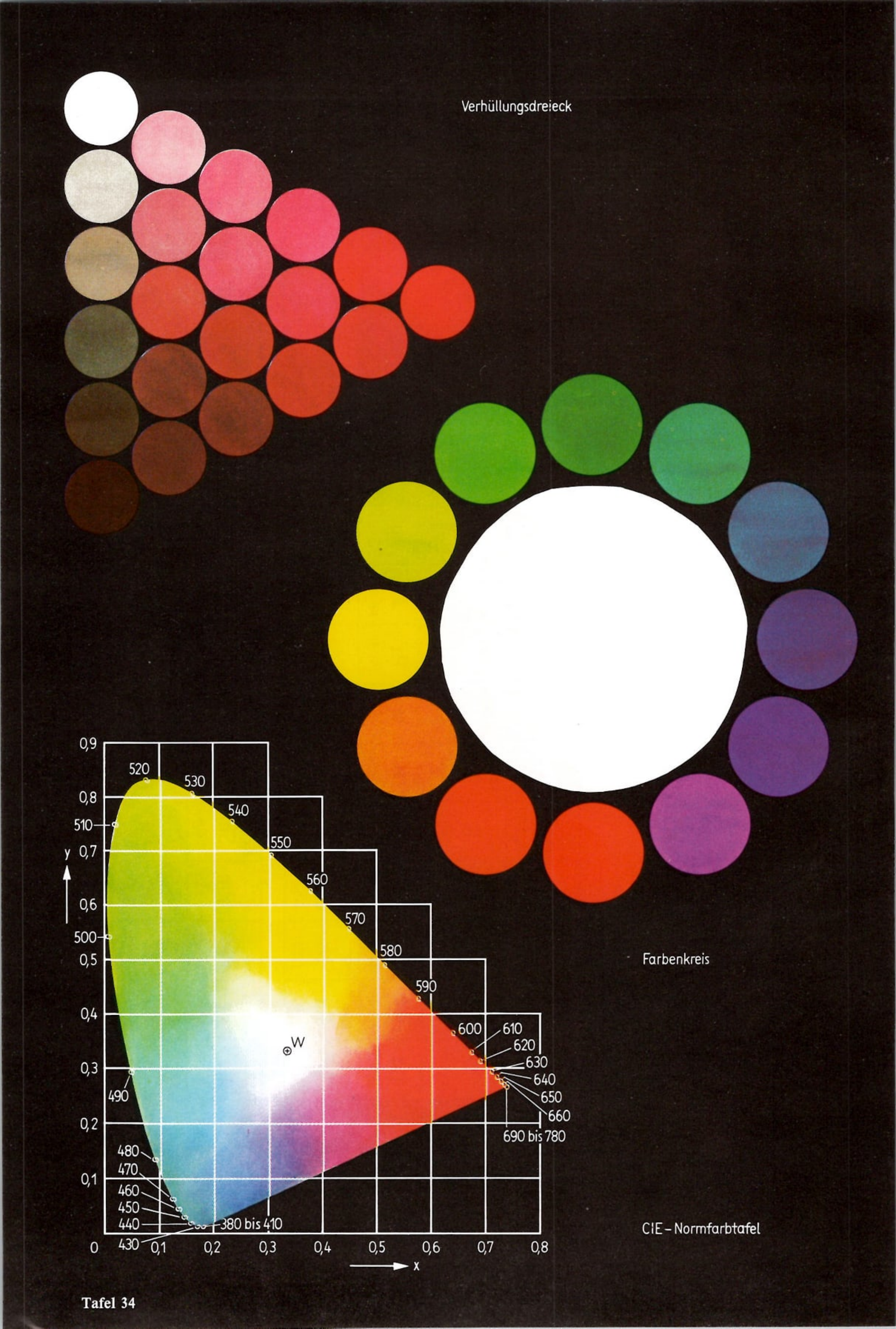
Ordnungszahl — 1
Elementsymbol — H
Atomgewicht[1] — 1,0079

1) Zahlen in Klammern jeweils für das Isotop mit der längsten bekannten Halbwertszeit

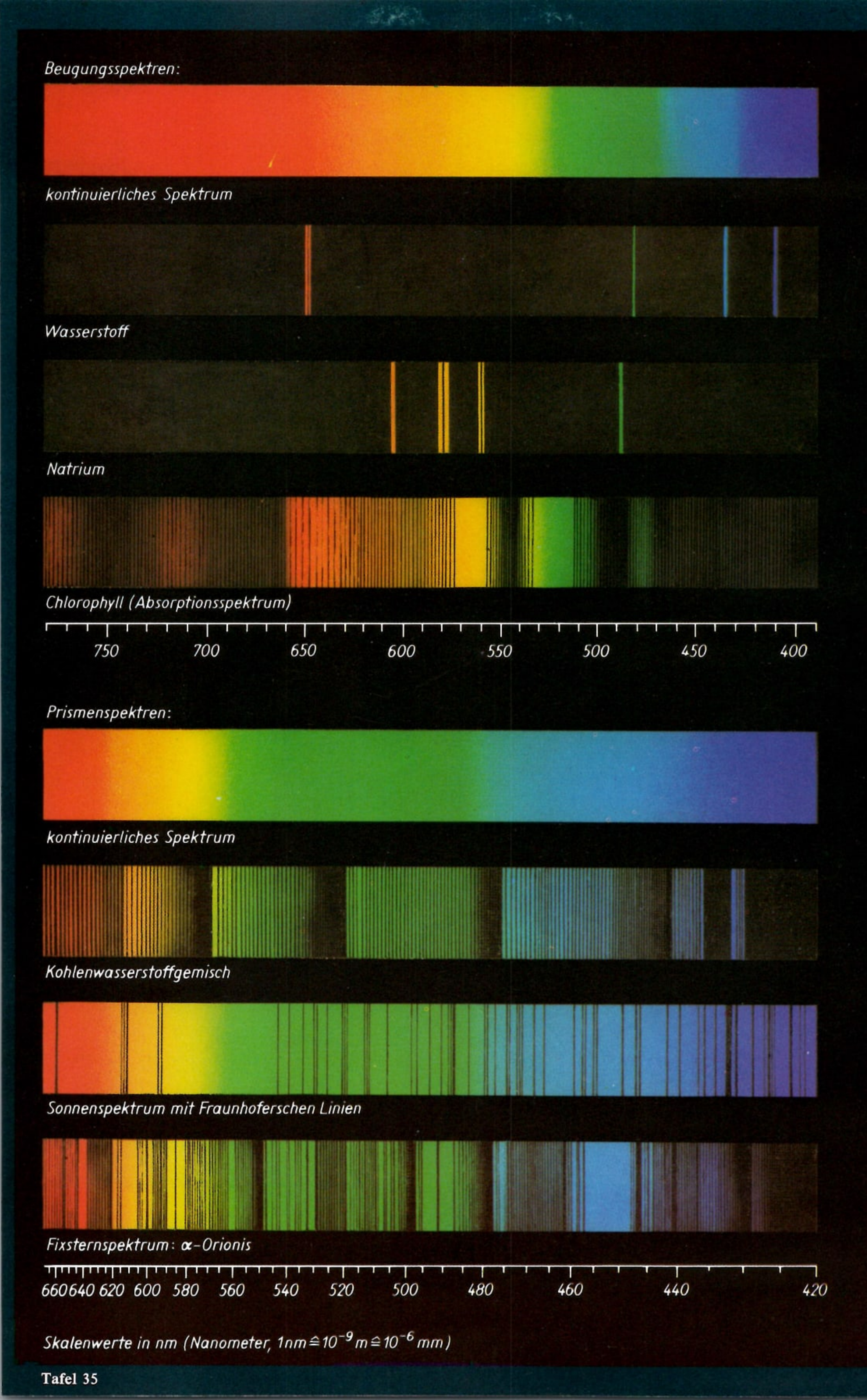
* Lanthanoiden-elemente	58 Ce 140,12	59 Pr 140,9077	60 Nd 144,24	61 Pm (145)	62 Sm 150,4	63 Eu 151,96	64 Gd 157,25	65 Tb 158,9254	66 Dy 162,50	67 Ho 164,9304	68 Er 167,26	69 Tm 168,9342	70 Yb 173,04	71 Lu 174,97
** Actinoiden-elemente	90 Th 232,0381	91 Pa 231,0359	92 U 238,029	93 Np 237,0482	94 Pu (244)	95 Am (243)	96 Cm (247)	97 Bk (247)	98 Cf (251)	99 Es (252)	100 Fm (257)	101 Md (258)	102 No (259)	103 Lr (260)

Hauptgruppenelemente
- Oxide, säurebildend
- Oxide, basenbildend
- Oxide, amphoter
- Edelgase

Nebengruppenelemente

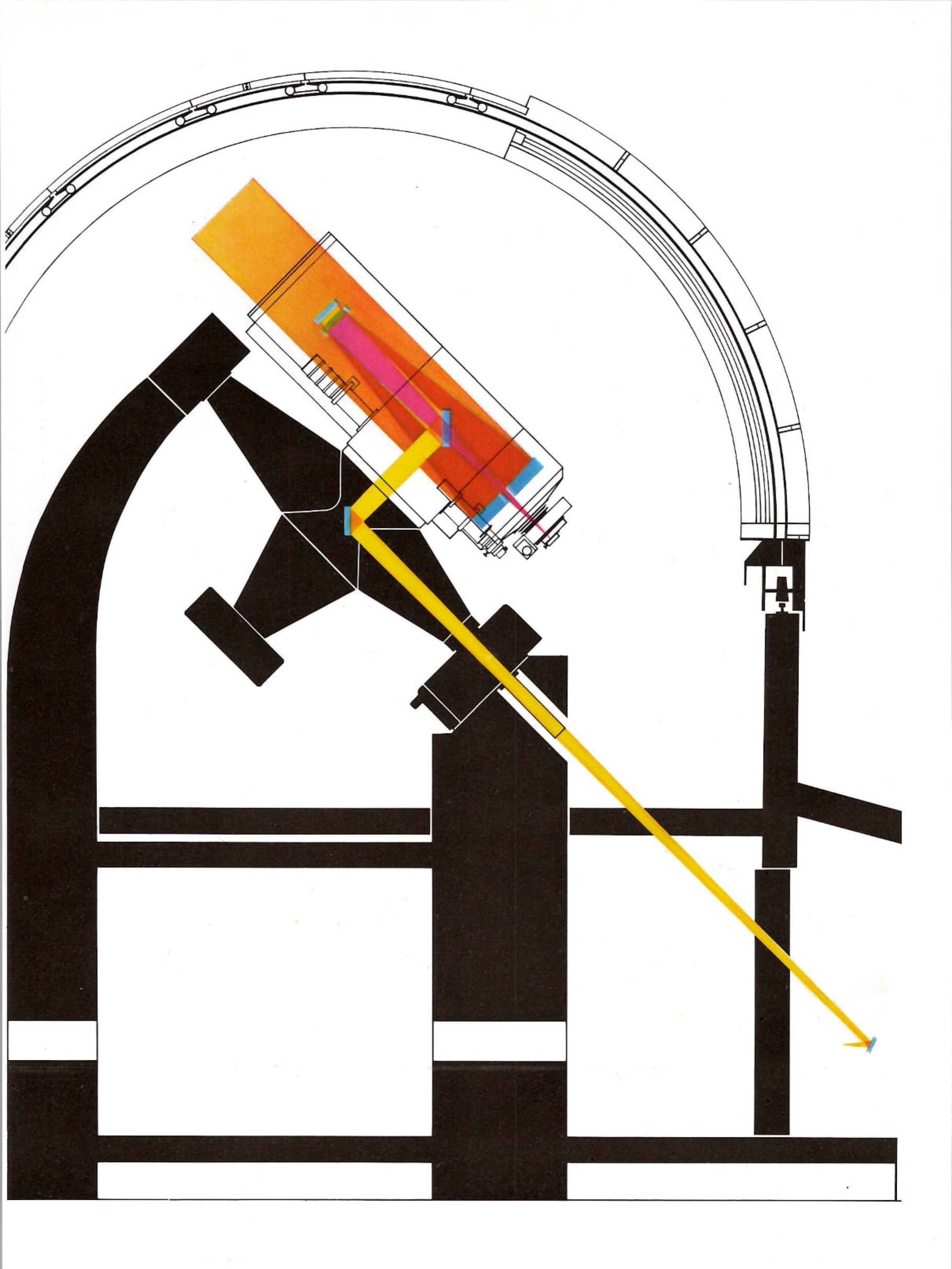


Verhüllungsdreieck
Farbenkreis
CIE-Normfarbtafel
0,9
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
y
x
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
520
530
540
550
560
570
580
590
600
610
620
630
640
650
660
690 bis 780
510
500
490
480
470
460
450
440
430
380 bis 410
W


Beugungsspektren:
kontinuierliches Spektrum
Wasserstoff
Natrium
Chlorophyll (Absorptionsspektrum)
750
700
650
600
550
500
450
400
Prismenspektren:
kontinuierliches Spektrum
Kohlenwasserstoffgemisch
Sonnenspektrum mit Fraunhoferschen Linien
Fixsternspektrum: α-Orionis
660
640
620
600
580
560
540
520
500
480
460
440
420
Skalenwerte in nm (Nanometer, 1nm ≙ 10⁻⁹ m ≙ 10⁻⁶ mm)



Tafel 36
1-m-Spiegelteleskop des VEB Carl Zeiss Jena



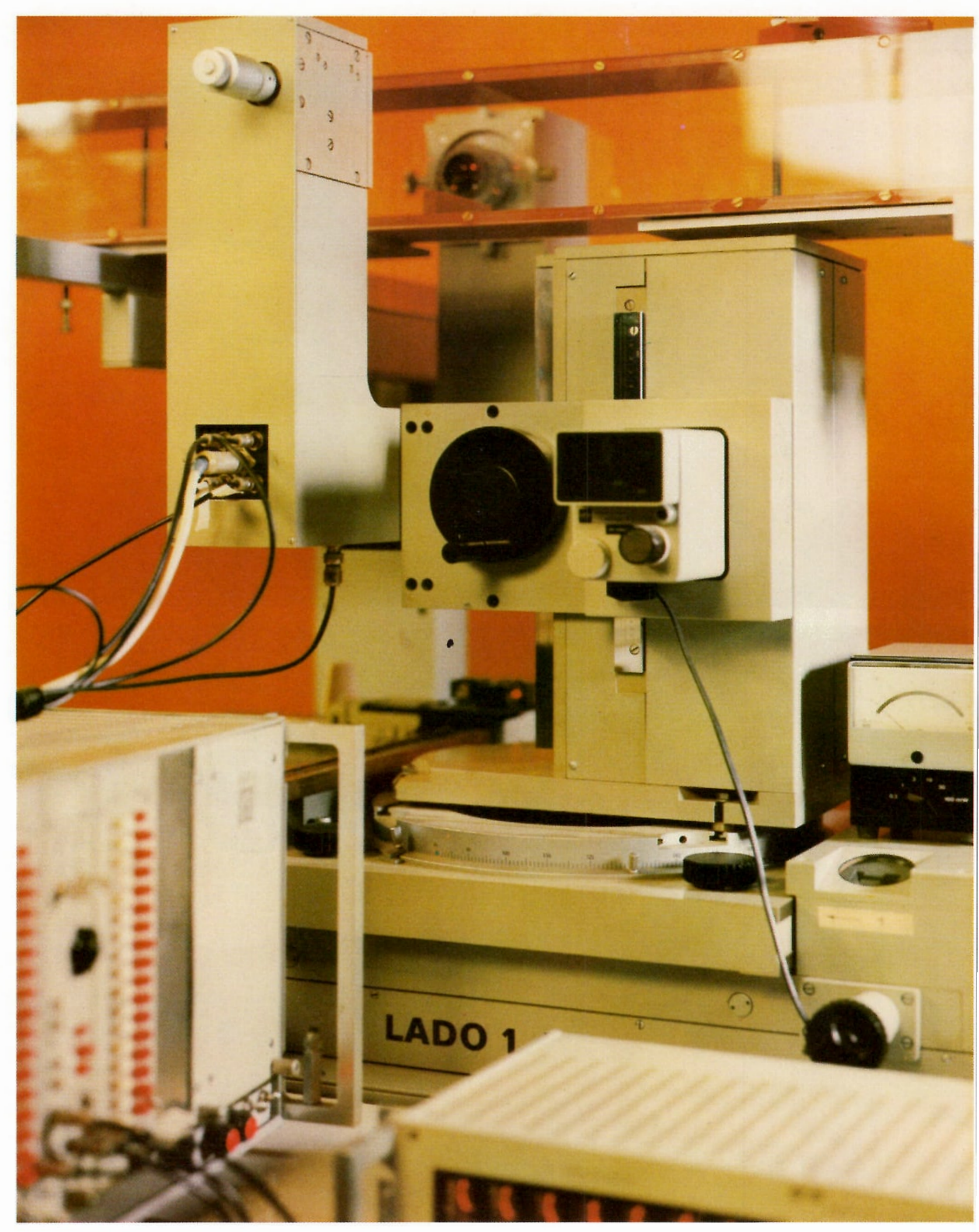
Tafel 37
Strahlengang in dem auf Tafel 36 dargestellten Spiegelteleskop



Tafel 38
Universaldurchlichtmikroskop „Jenaval“, die Vergrößerung ist zwischen 6,3- und 2000fach möglich. Die Modulatoren für positiven und negativen Phasenkontrast, differentiellen Interferenzkontrast und zentrales Dunkelfeld sind frei verfügbar.

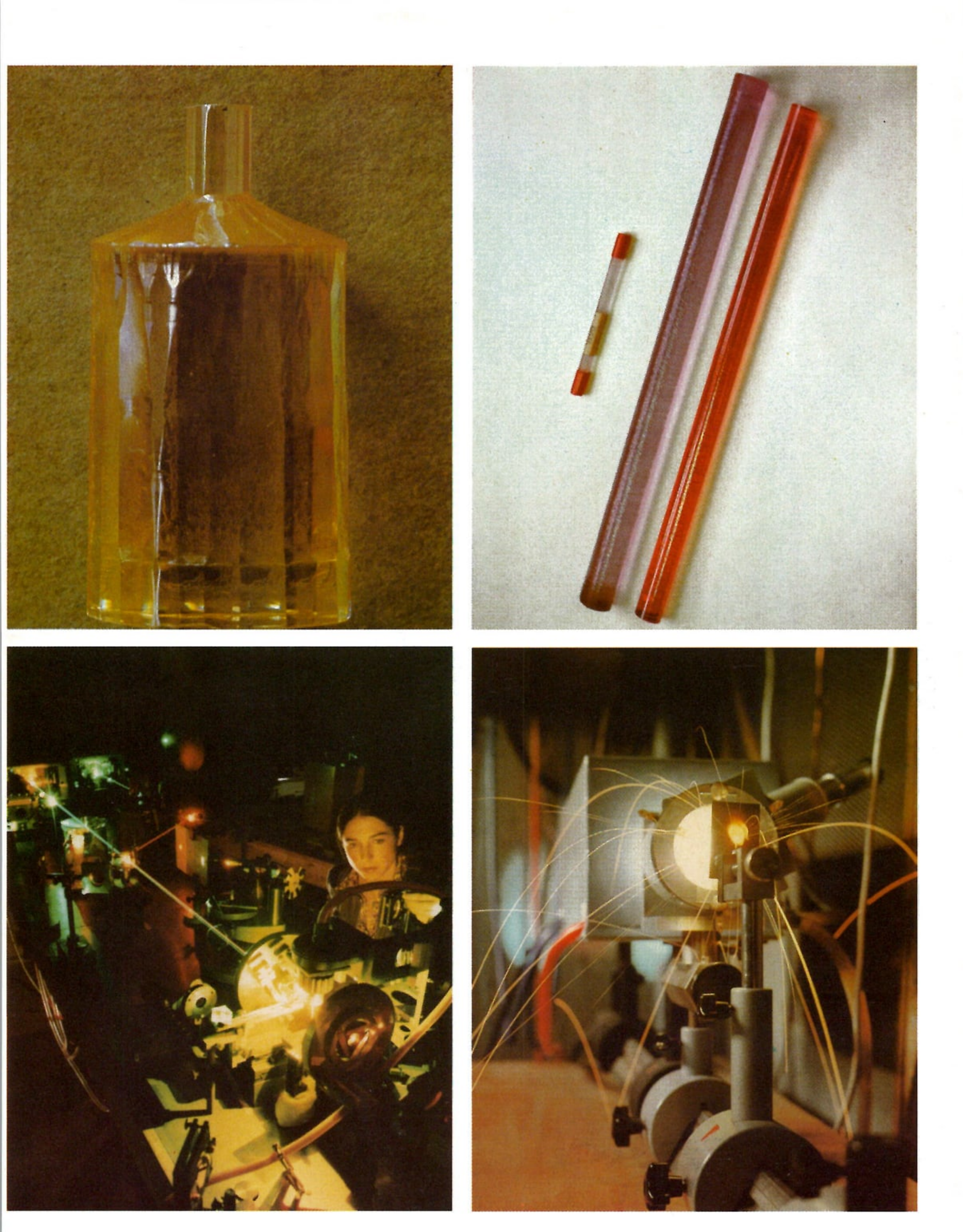
1
2

Tafel 39
1 Vollautomatisches registrierendes Spektralfotometer – 2 Laser-Mikrospektralanalysator für die Absorptions- und Emissionsspektralanalyse, im Hintergrund ein Plangitterspektograf.

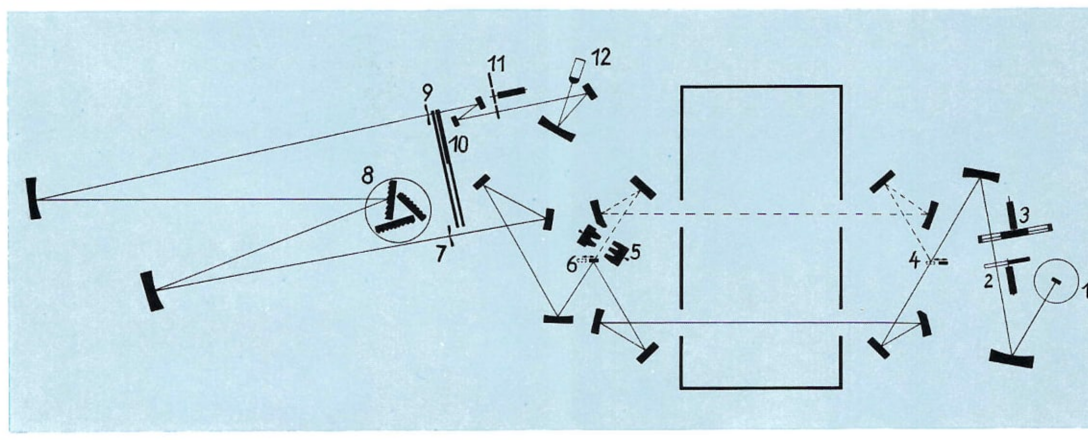


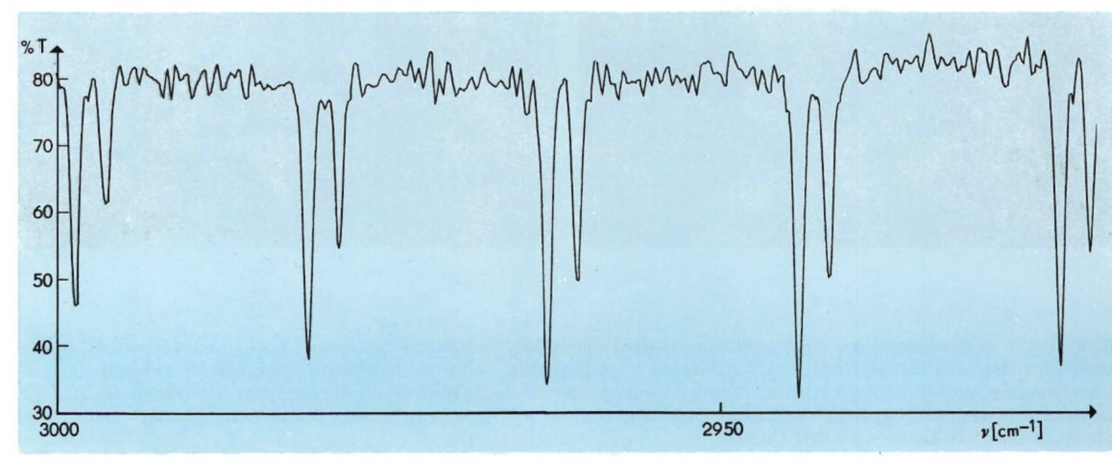
1 | 2 3
| 4 5

Tafel 40/41
1 Laser-Dopplergeschwindigkeitsmesser mit einem Meßbereich von 80 bis $8 \cdot 10^7$ m/s und einer Meßunsicherheit von 0,3 %



für die berührungslose Messung von Flüssigkeits- und Gasströmen – 2 In Lasergeräten verwendeter Schalt- und Modulatorkristall aus Strontium-Barium-Niobat – 3 Verschiedene Laser-Lagestäbe – 4 Durchstimmbarer Farbstofflaser. Durch optische Anregung von Farbstoffmolekülen mit einem Argonionenlaser kann die Wellenlänge der Strahlung in weiten Bereichen des sichtbaren und infraroten Spektralbereiches variiert werden – 5 Aufschmelzen eines Schamottesteines durch die Strahlung eines CO_2-Lasers mit 200 W Leistung.

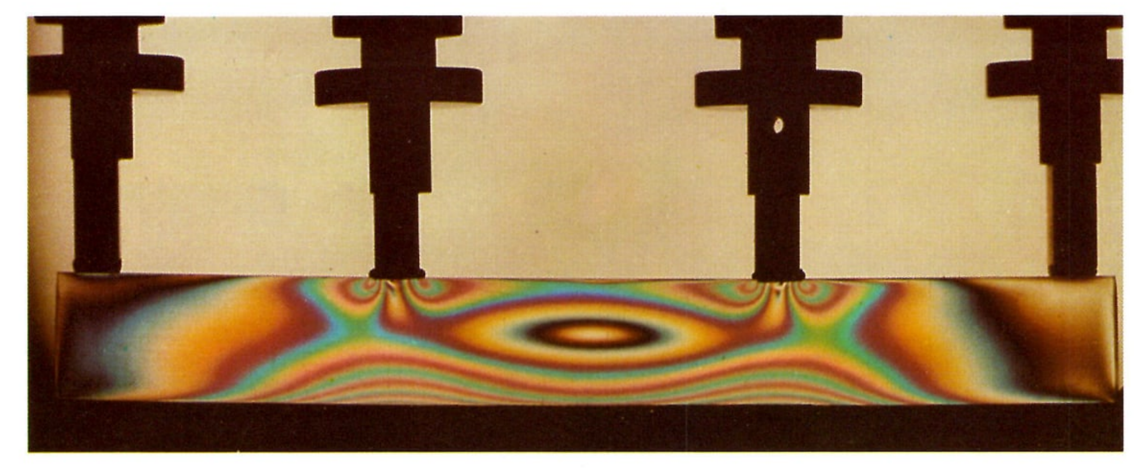

1
2
3
4
5
6
7
8
9
10
11
12


% T
80
70
60
50
40
30
3000
2950
ν [cm⁻¹]

1 | 4
2
3 | 5

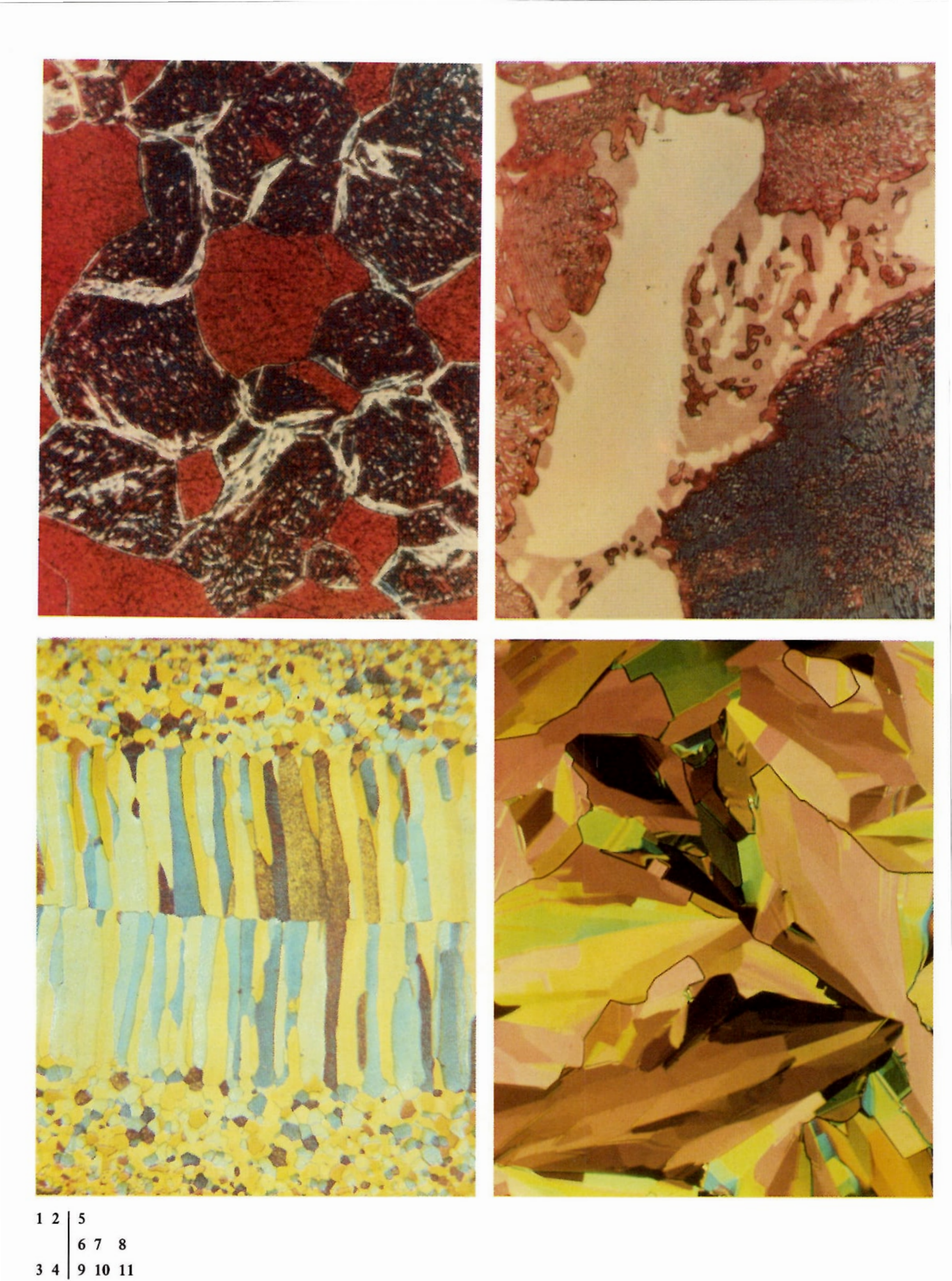
Tafeln 42 und 43

1 Hochautomatisierte Infrarot-Spektralfotometer für den Bereich von 400 bis 200 cm^{-1} – **2** Optisches Schema (Strahlengang) des Spektralfotometers; 1 Strahlungsquelle, 2 Modulator, 3 Polarisator, 4 und 6 Schwingspiegel, 5 Vergleichsstrahlschwächer, 7 Eintrittsspalt, 8 Beugungsgitter, 9 Austrittsspalt, 10 Ordnungsfilter, 11 Filter, 12 Thermoelement – **3** Spektrum von Chlorwasserstoff – **4** Atom-Absorptions-Spektralfotometer für den Spektralbereich von 190 bis 865 nm. Die Gerätesteuerung sowie die Signalerfassung, -verarbeitung und -auswertung erfolgen im integrierten Mikrorechner – **5** Strahlengang des Spektralfotometers: 1 Linienstrahler, 2/3 Kontinuumstrahler, 4 Chopper, 5 Atomisator, 6 Recombiner, 7 Gitter, 8 Filter, 9 Empfänger.



1 | 3
2 | 4

Tafeln 44 und 45
1 Spannungsoptische Aufnahme eines Fundamentbalkens unter Druckbelastung – 2 Spannungsoptische Aufnahme der Zwischenwand eines Generatorgehäuses – 3 Kristallziehen im Schwebezonenverfahren. Die HF-Spule heizt den Siliciumstab in einem schmalen Bereich hoch auf – 4 Auswahl synthetischer Einkristalle, die hauptsächlich nach dem Čzochralski-Verfahren in dem Spezialbetrieb für einkristalline Werkstoffe „Cristalec" des Forschungszentrums CNRS Grenoble hergestellt worden sind; *oben links* Halbleiterkristalle von Germanium und Silicium, *darunter* bzw. *rechts* Einkristalle verschiedener Metalle, in der *Mitte* rote Stäbe für Rubinlaser, *links unten* schön gefärbte künstliche Halbedelsteine, *unten rechts* Einkristalle magnetischer Materialien



1	2	5		
		6	7	8
3	4	9	10	11

Tafeln 46 und 47

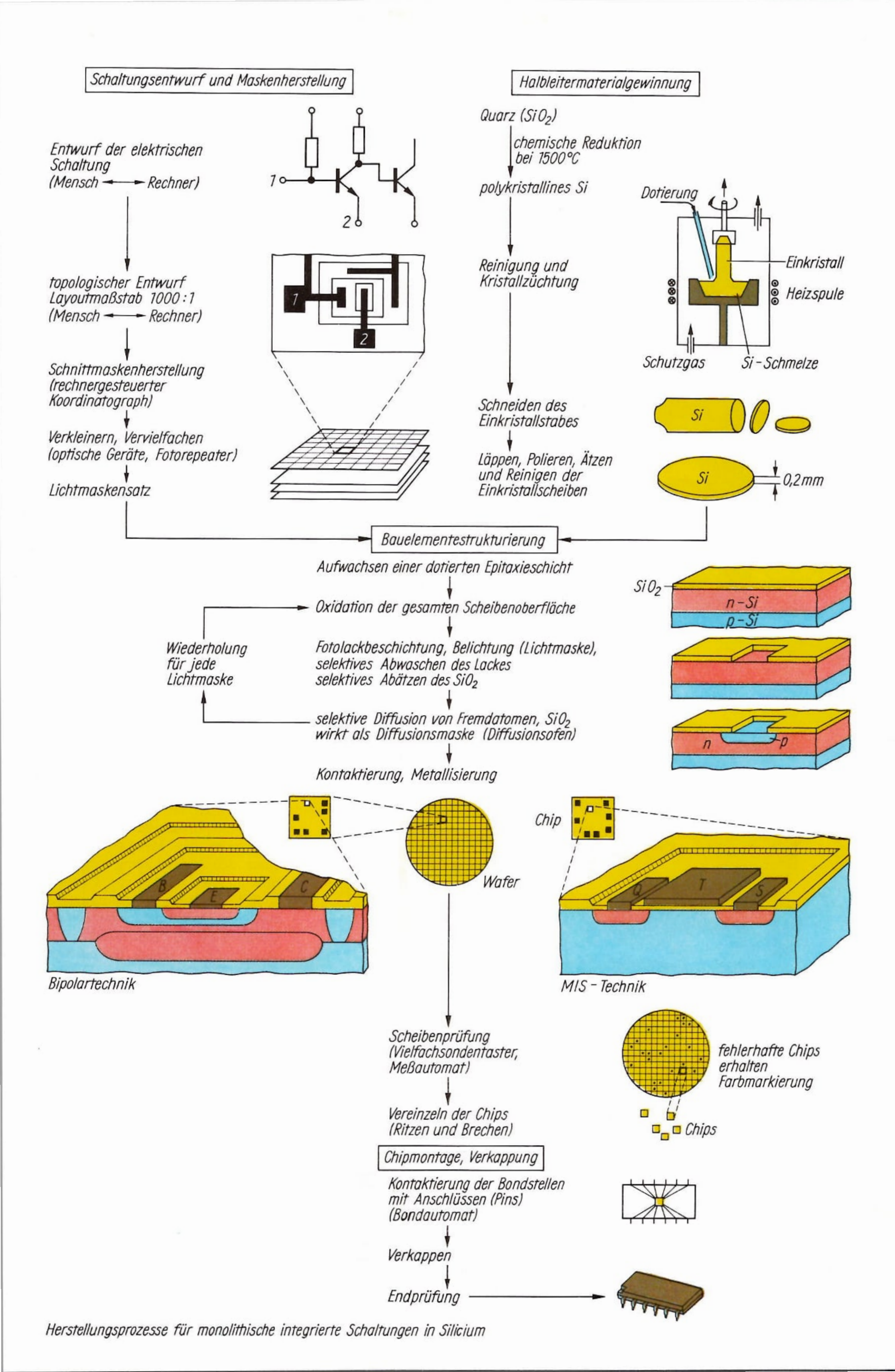
1 Stahl mit 0,2 % C, 17 % Cr, 1 % Mo von 1 200 °C in Öl abgeschreckt und 5 min lang bei 500 °C in Luft oxidiert; *rote Grundmasse* Ferrit *weiße Kristalle* Martensit (500fache Vergrößerung) – **2** Gußeisen 5 min lang bei 300 °C an Luft oxidiert; Eisen-

carbid *weiß*, Ferrit *blau* und *braun*, Eisenphosphid *graubraun* – **3** Typisches Erstarrungsgefüge (Molybdän, geätzt) – **4** Smektischer Flüssigkristall, G-Phasentyp, Mosaiktextur 3-n-Hexyl-6-[4-n-pentyloxyphenyl]-pyridazin (83 °C, 50fache Vergrößerung) – **5** Scanning-Elektronenmikroskop mit Mikroanalysezusatz – **6–11** Mikroanalyse einer mineralogischen Probe – **6** Darstellung des gesamten Röntgen-Emissionsspektrums der Probe – **7** Aufspreizung eines Teiles des Spektrums durch den angeschlossenen Rechner – **8** Abbildung charakteristischer Aufnahmedaten auf dem Computerbildschirm – **9** Vom Rechner ermittelter und zu subtrahierender Bremsstrahlungsuntergrund (weiß markiert) – **10** Die vom Computer diagnostizierten Röntgen-Linien der wichtigsten Elemente – **11** Überlagerung von Röntgen-Linien von Schwefel S und Blei Pb im gelb dargestellten Bereich. Eine Trennung ist durch den Computer möglich.

1 | 2 | 3

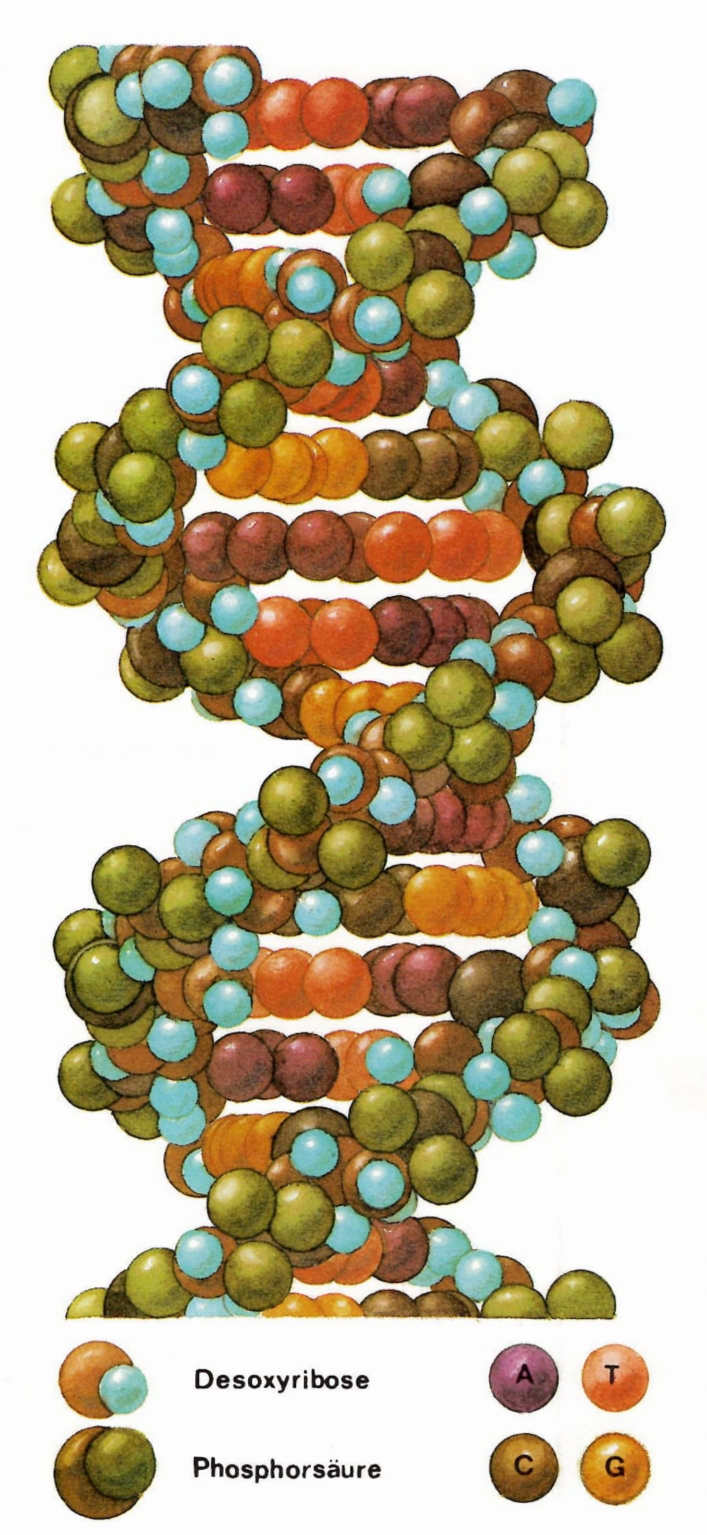
Tafeln 48 und 49

1 Röntgen-Zählrohrdiffraktometer mit aufgesetztem Zusatz für Texturuntersuchungen sowie Steuer- und Strahlenmeßelektronik – 2 Durchstrahlungsmikroskop mit bis zu 250000facher Vergrößerung. Die Auflösung von Netzebenenabständen ist bis zu 0,2 nm möglich – 3 Herstellungsschritte für monolithisch integrierte Schaltungen in Silicium.



Herstellungsprozesse für monolithische integrierte Schaltungen in Silicium

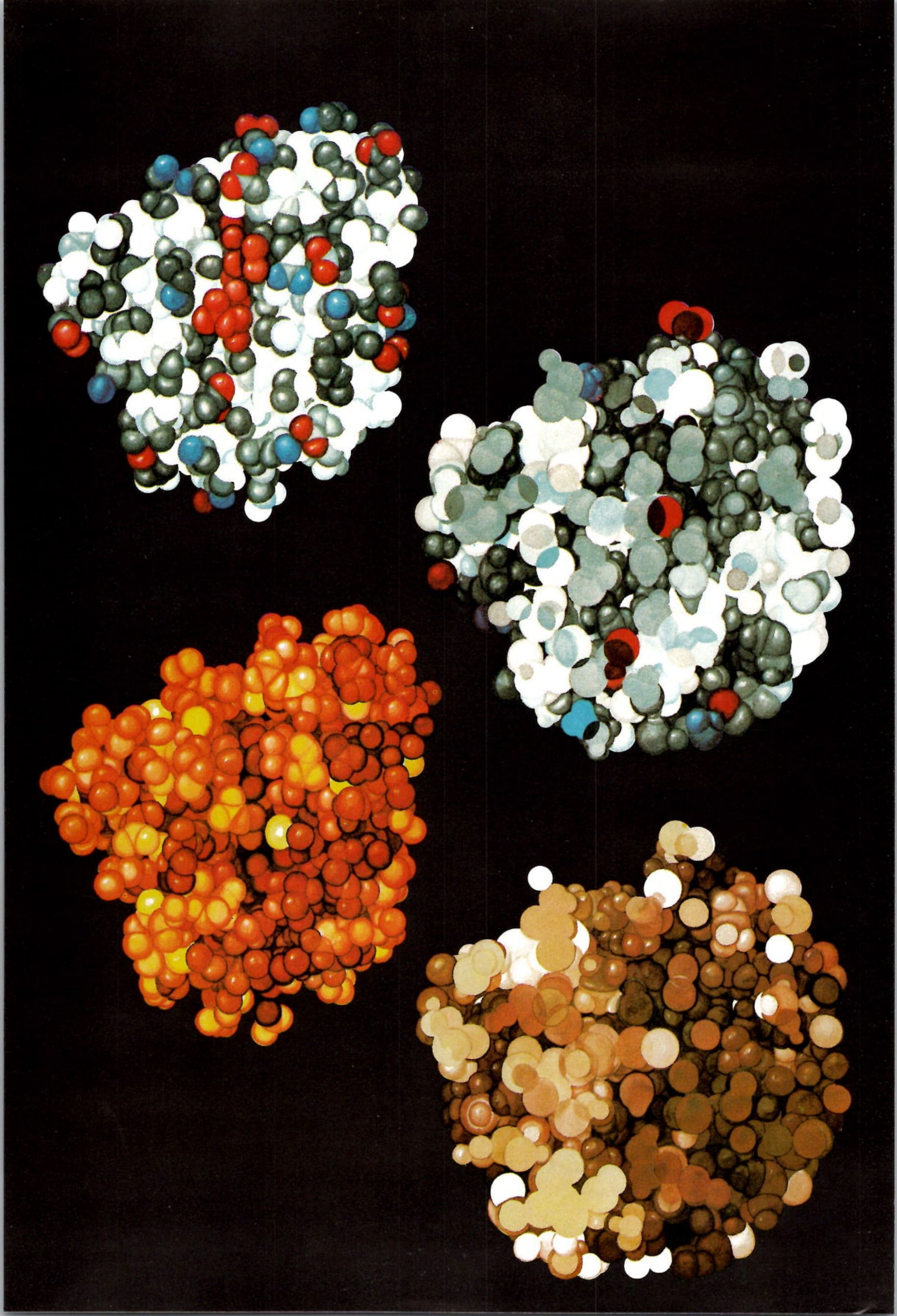
1 | 2

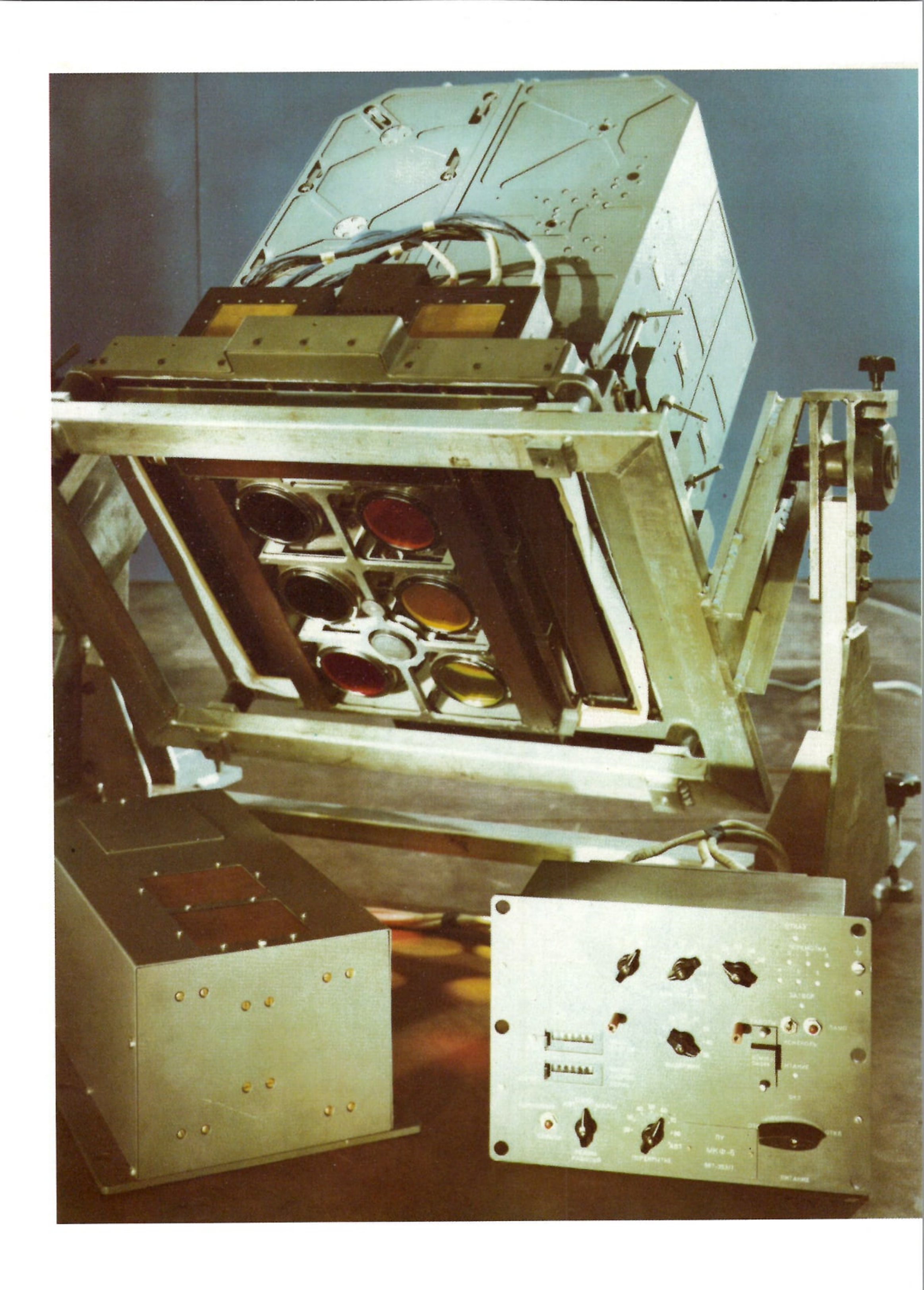


Tafel 50 und 51

1 Kugelmodell der Doppelhelix

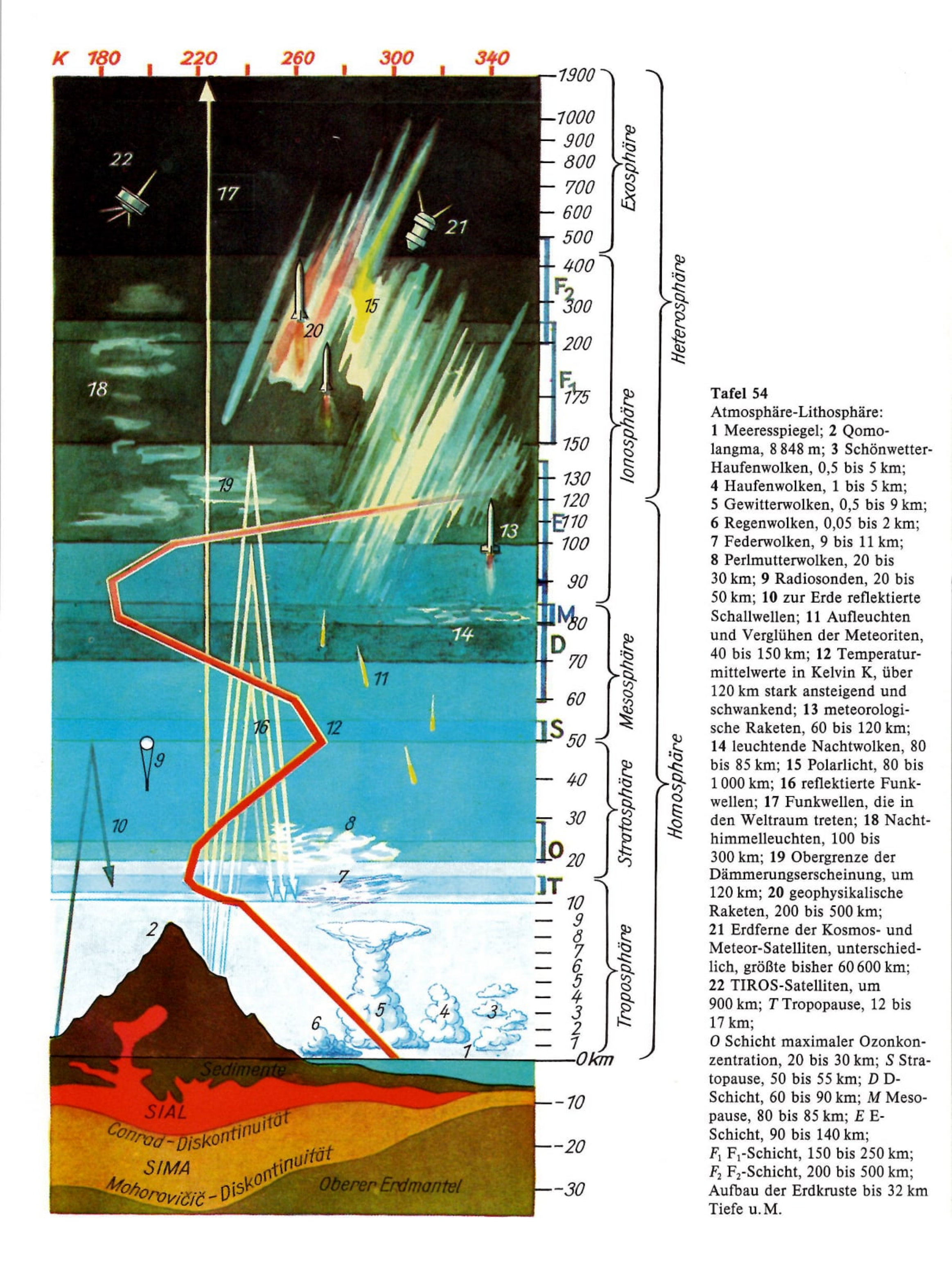
2 Darstellung des Proteins Myoglobin vermittels der sogenannten „Computer Graphics" nach R. J. FELDMANN, *links oben* Darstellung der Ladungsverteilung auf der Oberfläche des Moleküls, wobei rot negativ geladene, blau positiv geladene, dunkelgrau hydrophobe und weiß hydrophile Atome kennzeichnen. *Rechts oben* zeigt das Myoglobin-Molekül in einer Schnittdarstellung senkrecht zur funktionellen Hämgruppe. Deutlich wird auch hier wieder die bevorzugte Anordnung von geladenen und hydrophilen Gruppen auf der Außenseite und von hydrophoben im Innern (sogenannte „hydrophobe Tasche" um die Hämgruppe). *Links unten* Beweglichkeit der Myoglobin-Atome (nach Röntgen-Streuungsuntersuchungen von H. FRAUENFELDER u. a.). Dunkle Färbung entspricht geringer Beweglichkeit (kleines mittleres Schwankungsquadrat), helle hoher Beweglichkeit. *Rechts unten* zeigt die Beweglichkeit wiederum in der Schnittdarstellung (wie in b). Offenbar ist der rechte Teil, wo Sauerstoff gebunden wird, weniger beweglich als der linke, „halbflüssige" Bereich, der für den schnellen Transport des Sauerstoffs durch „zweidimensionale Diffusion" zur Hämgruppe von wesentlicher Bedeutung sein dürfte.





Tafel 52
Multispektralkamera MKF-6 für die Fernerkundung der Erde mit Hilfe von Satelliten.

Tafel 53
Aufnahmen der Multispektralkamera MKF-6, auf denen unterschiedliche Vegetationsgebiete erkennbar sind.



Tafel 54
Atmosphäre-Lithosphäre:
1 Meeresspiegel; **2** Qomolangma, 8 848 m; **3** Schönwetter-Haufenwolken, 0,5 bis 5 km; **4** Haufenwolken, 1 bis 5 km; **5** Gewitterwolken, 0,5 bis 9 km; **6** Regenwolken, 0,05 bis 2 km; **7** Federwolken, 9 bis 11 km; **8** Perlmutterwolken, 20 bis 30 km; **9** Radiosonden, 20 bis 50 km; **10** zur Erde reflektierte Schallwellen; **11** Aufleuchten und Verglühen der Meteoriten, 40 bis 150 km; **12** Temperaturmittelwerte in Kelvin K, über 120 km stark ansteigend und schwankend; **13** meteorologische Raketen, 60 bis 120 km; **14** leuchtende Nachtwolken, 80 bis 85 km; **15** Polarlicht, 80 bis 1 000 km; **16** reflektierte Funkwellen; **17** Funkwellen, die in den Weltraum treten; **18** Nachthimmelleuchten, 100 bis 300 km; **19** Obergrenze der Dämmerungserscheinung, um 120 km; **20** geophysikalische Raketen, 200 bis 500 km; **21** Erdferne der Kosmos- und Meteor-Satelliten, unterschiedlich, größte bisher 60 600 km; **22** TIROS-Satelliten, um 900 km; *T* Tropopause, 12 bis 17 km;
O Schicht maximaler Ozonkonzentration, 20 bis 30 km; *S* Stratopause, 50 bis 55 km; *D* D-Schicht, 60 bis 90 km; *M* Mesopause, 80 bis 85 km; *E* E-Schicht, 90 bis 140 km;
F_1 F_1-Schicht, 150 bis 250 km;
F_2 F_2-Schicht, 200 bis 500 km;
Aufbau der Erdkruste bis 32 km Tiefe u. M.

1
2

Tafel 55
1 Kräftiges Polarlicht mit Mondaufgang über einer Beobachtungsstation in der Ostantarktis – 2 Rote Korona vom 13. Juli 1982 über der Antarktis







Tafel 56

1 Aufnahme des Saturn von der Voyager-1-Sonde aus 34 Mill. km Entfernung im ultravioletten Licht. Die im UV-Licht hellen Zonen erscheinen in diesem Falschfarbenfoto türkis bis violett, die dunklen orange – **2** Aufnahme des Saturns von der Voyager 1 aus im sichtbaren Licht (von oben, d. h. sonnenzugewandte Seite) – **3** Ringsystem des Saturns von unten (sonnenabgewandte Seite). Der schmale Streifen außen ist der F-Ring, gefolgt vom A-Ring mit Encke-Teilung. Die Cassini-Teilung zwischen A- und B-Ring (hell) ist strukturiert und ebenso wie der innerhalb des B-Ringes liegende C-Ring nicht leer. Die Encke-Teilung scheint wirklich leer zu sein.

Kleine Enzyklopädie
Physik

Kleine Enzyklopädie Physik

Wellen
Teilchen
Strahlung
Energie

Raum
Umwelt
Erde
Sonne
Weltall

Zeit
Relativität
Urknall
Rotverschiebung

Elektrizität
Magnetismus
Temperatur
Entropie

Quarks
Leptonen
Atome
Moleküle
Leben

